SECRETED PROTEINS

TECHNICAL FIELD

The invention relates to novel nucleic acids, secreted proteins encoded by these nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, developmental, and metabolic disorders. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and secreted proteins.

BACKGROUND OF THE INVENTION

Protein transport and secretion are essential for cellular function. Protein transport is mediated by a signal peptide located at the amino terminus of the protein to be transported or secreted. The signal peptide is comprised of about ten to twenty hydrophobic amino acids which target the nascent protein from the ribosome to a particular membrane bound compartment such as the endoplasmic reticulum (ER). Proteins targeted to the ER may either proceed through the secretory pathway or remain in any of the secretory organelles such as the ER, Golgi apparatus, or lysosomes. Proteins that transit through the secretory pathway are either secreted into the extracellular space or retained in the plasma membrane. Proteins that are retained in the plasma membrane contain one or more transmembrane domains, each comprised of about 20 hydrophobic amino acid residues. Secreted proteins are generally synthesized as inactive precursors that are activated by post-translational processing events during transit through the secretory pathway. Such events include glycosylation, proteolysis, and removal of the signal peptide by a signal peptidase. Other events that may occur during protein transport include chaperone-dependent unfolding and folding of the nascent protein and interaction of the protein with a receptor or pore complex. Examples of secreted proteins with amino terminal signal peptides are discussed below and include proteins with important roles in cell-to-cell signaling. Such proteins include transmembrane receptors and cell surface markers, extracellular matrix molecules, cytokines, hormones, growth and differentiation factors, enzymes, neuropeptides, vasomediators, cell surface markers, and antigen recognition molecules. (Reviewed in Alberts, B. et al. (1994) Molecular Biology of The Cell, Garland Publishing, New York, NY, pp. 557-560, 582-592.)

Cell surface markers include cell surface antigens identified on leukocytic cells of the immune system. These antigens have been identified using systematic, monoclonal antibody (mAb)-based "shot gun" techniques. These techniques have resulted in the production of hundreds of mAbs directed against unknown cell surface leukocytic antigens. These antigens have been grouped into "clusters of differentiation" based on common immunocytochemical localization patterns in various

differentiated and undifferentiated leukocytic cell types. Antigens in a given cluster are presumed to identify a single cell surface protein and are assigned a "cluster of differentiation" or "CD" designation. Some of the genes encoding proteins identified by CD antigens have been cloned and verified by standard molecular biology techniques. CD antigens have been characterized as both transmembrane proteins and cell surface proteins anchored to the plasma membrane via covalent attachment to fatty acid-containing glycolipids such as glycosylphosphatidylinositol (GPI). (Reviewed in Barclay, A.N. et al. (1995) The Leucocyte Antigen Facts Book, Academic Press, San Diego, CA, pp. 17-20.)

Matrix proteins (MPs) are transmembrane and extracellular proteins which function in formation, growth, remodeling, and maintenance of tissues and as important mediators and regulators of the inflammatory response. The expression and balance of MPs may be perturbed by biochemical changes that result from congenital, epigenetic, or infectious diseases. In addition, MPs affect leukocyte migration, proliferation, differentiation, and activation in the immune response. MPs are frequently characterized by the presence of one or more domains which may include collagen-like domains, EGF-like domains, immunoglobulin-like domains, and fibronectin-like domains. In addition, MPs may be heavily glycosylated and may contain an Arginine-Glycine-Aspartate (RGD) tripeptide motif which may play a role in adhesive interactions. MPs include extracellular proteins such as fibronectin, collagen, galectin, vitronectin and its proteolytic derivative somatomedin B; and cell adhesion receptors such as cell adhesion molecules (CAMs), cadherins, and integrins. (Reviewed in Ayad, S. et al. (1994) The Extracellular Matrix Facts Book, Academic Press, San Diego, CA, pp. 2-16; Ruoslahti, E. (1997) Kidney Int. 51:1413-1417; Sjaastad, M.D. and W.J. Nelson (1997) BioEssays 19:47-55.)

Mucins are highly glycosylated glycoproteins that are the major structural component of the mucus gel. The physiological functions of mucins are cytoprotection, mechanical protection, maintenance of viscosity in secretions, and cellular recognition. MUC6 is a human gastric mucin that is also found in gall bladder, pancreas, seminal vesicles, and female reproductive tract (Toribara, N.W. et al. (1997) J. Biol. Chem. 272:16398-16403). The MUC6 gene has been mapped to human chromosome 11 (Toribara, N.W. et al. (1993) J. Biol. Chem. 268:5879-5885). Hemomucin is a novel *Drosophila* surface mucin that may be involved in the induction of antibacterial effector molecules (Theopold, U. et al. (1996) J. Biol. Chem. 217:12708-12715).

Tuftelins are one of four different enamel matrix proteins that have been identified so far. The other three known enamel matrix proteins are the amelogenins, enamelin and ameloblastin. Assembly of the enamel extracellular matrix from these component proteins is believed to be critical in producing a matrix competent to undergo mineral replacement (Paine, C.T. et al. (1998) Connect Tissue Res. 38:257-267). Tuftelin mRNA has been found to be expressed in human ameloblastoma

tumor, a non-mineralized odontogenic tumor (Deutsch, D. et al. (1998) Connect. Tissue Res. 39:177-184).

Olfactomedin-related proteins are extracellular matrix, secreted glycoproteins with conserved C-terminal motifs. They are expressed in a wide variety of tissues and in a broad range of species, from *Caenorhabditis elegans* to *Homo sapiens*. Olfactomedin-related proteins comprise a gene family with at least 5 family members in humans. One of the five, TIGR/myocilin protein, is expressed in the eye and is associated with the pathogenesis of glaucoma (Kulkarni, N.H. et al. (2000) Genet. Res. 76:41-50). Research by Yokoyama, M. et al. (1996; DNA Res. 3:311-320) found a 135-amino acid protein, termed AMY, having 96% sequence identity with rat neuronal olfactomedin-releated ER localized protein in a neuroblastoma cell line cDNA library, suggesting an essential role for AMY in nerve tissue. Neuron-specific olfactomedin-related glycoproteins isolated from rat brain cDNA libraries show strong sequence similarity with olfactomedin. This similarity is suggestive of a matrix-related function of these glycoproteins in neurons and neurosecretory cells (Danielson, P.E. et al. (1994) J. Neurosci. Res. 38:468-478).

Mac-2 binding protein is a 90-kD serum protein (90K), a secreted glycoprotein isolated from both the human breast carcinoma cell line SK-BR-3, and human breast milk. It specifically binds to a human macrophage-associated lectin, Mac-2. Structurally, the mature protein is 567 amino acids in length and is proceeded by an 18-amino acid leader. There are 16 cysteines and seven potential N-linked glycosylation sites. The first 106 amino acids represent a domain very similar to an ancient protein superfamily defined by a macrophage scavenger receptor cysteine-rich domain (Koths, K. et al. (1993) J. Biol. Chem. 268:14245-14249). 90K is elevated in the serum of subpopulations of AIDS patients and is expressed at varying levels in primary tumor samples and tumor cell lines. Ullrich, A. et al. (1994; J. Biol. Chem. 269:18401-18407) have demonstrated that 90K stimulates host defense systems and can induce interleukin-2 secretion. This immune stimulation is proposed to be a result of oncogenic transformation, viral infection or pathogenic invasion (Ullrich et al., *supra*).

Semaphorins are a large group of axonal guidance molecules consisting of at least 30 different members and are found in vertebrates, invertebrates, and even certain viruses. All semaphorins contain the sema domain which is approximately 500 amino acids in length. Neuropilin, a semaphorin receptor, has been shown to promote neurite outgrowth *in vitro*. The extracellular region of neuropilins consists of three different domains: CUB, discoidin, and MAM domains. The CUB and the MAM motifs of neuropilin have been suggested to have roles in protein-protein interactions and are thought to be involved in the binding of semaphorins through the sema and the C-terminal domains (reviewed in Raper, J.A. (2000) Curr. Opin. Neurobiol. 10:88-94). Plexins are neuronal cell surface molecules that mediate cell adhesion via a homophilic binding mechanism in the presence of calcium ions. Plexins have been shown to be expressed in the receptors and neurons of

particular sensory systems (Ohta, K. et al. (1995) Cell 14:1189-1199). There is evidence that suggests that some plexins function to control motor and CNS axon guidance in the developing nervous system. Plexins, which themselves contain complete semaphorin domains, may be both the ancestors of classical semaphorins and binding partners for semaphorins (Winberg, M.L. et al (1998) Cell 95:903-916).

Human pregnancy-specific beta 1-glycoprotein (PSG) is a family of closely related glycoproteins of molecular weights of 72 KDa, 64KDa, 62KDa, and 54KDa. Together with the carcinoembryonic antigen, they comprise a subfamily within the immunoglobulin superfamily (Plouzek, C.A. and J.Y. Chou, (1991) Endocrinology 129:950-958) Different subpopulations of PSG have been found to be produced by the trophoblasts of the human placenta, and the amnionic and chorionic membranes (Plouzek, C.A. et al. (1993) Placenta 14:277-285).

Torsion dystonia is an autosomal dominant movement disorder consisting of involuntary muscular contractions. The disorder has been linked to a 3-base pair mutation in the DYT-1 gene, which encodes torsin A (Ozelius, L.J. et al. (1997) Nat. Genet. 17:40-48). Torsin A bears significant homology to the Hsp100/Clp family of ATPase chaperones, which are conserved in humans, rats, mice, and *C. elegans*. Strong expression of DYT-1 in neuronal processes indicates a potential role for torsins in synaptic communication (Kustedjo, K. et al. (2000) J. Biol. Chem. 275:27933-27939 and Konakova M. et al. (2001) Arch. Neurol. 58:921-927).

Autocrine motility factor (AMF) is one of the motility cytokines regulating tumor cell migration; therefore identification of the signaling pathway coupled with it has critical importance. Autocrine motility factor receptor (AMFR) expression has been found to be associated with tumor progression in thymoma (Ohta Y. et al. (2000) Int. J. Oncol. 17:259-264). AMFR is a cell surface glycoprotein of molecular weight 78KDa.

Hormones are secreted molecules that travel through the circulation and bind to specific receptors on the surface of, or within, target cells. Although they have diverse biochemical compositions and mechanisms of action, hormones can be grouped into two categories. One category includes small lipophilic hormones that diffuse through the plasma membrane of target cells, bind to cytosolic or nuclear receptors, and form a complex that alters gene expression. Examples of these molecules include retinoic acid, thyroxine, and the cholesterol-derived steroid hormones such as progesterone, estrogen, testosterone, cortisol, and aldosterone. The second category includes hydrophilic hormones that function by binding to cell surface receptors that transduce signals across the plasma membrane. Examples of such hormones include amino acid derivatives such as catecholamines (epinephrine, norepinephrine) and histamine, and peptide hormones such as glucagon, insulin, gastrin, secretin, cholecystokinin, adrenocorticotropic hormone, follicle stimulating hormone, luteinizing hormone, thyroid stimulating hormone, and vasopressin. (See, for example, Lodish et al.

(1995) Molecular Cell Biology, Scientific American Books Inc., New York, NY, pp. 856-864.)

Pro-opiomelanocortin (POMC) is the precursor polypeptide of corticotropin (ACTH), a hormone synthesized by the anterior pituitary gland, which functions in the stimulation of the adrenal cortex. POMC is also the precursor polypeptide of the hormone beta-lipotropin (beta-LPH). Each hormone includes smaller peptides with distinct biological activities: alpha-melanotropin (alpha-MSH) and corticotropin-like intermediate lobe peptide (CLIP) are formed from ACTH; gamma-lipotropin (gamma-LPH) and beta-endorphin are peptide components of beta-LPH; while beta-MSH is contained within gamma-LPH. Adrenal insufficiency due to ACTH deficiency, resulting from a genetic mutation in exons 2 and 3 of POMC results in an endocrine disorder characterized by early-onset obesity, adrenal insufficiency, and red hair pigmentation (Chretien, M. et al. (1979) Can. J. Biochem. 57:1111-1121; Krude, H. et al. (1998) Nat. Genet. 19:155-157; Online Mendelian Inheritance in Man (OMIM) 176830).

Growth and differentiation factors are secreted proteins which function in intercellular communication. Some factors require oligomerization or association with membrane proteins for activity. Complex interactions among these factors and their receptors trigger intracellular signal transduction pathways that stimulate or inhibit cell division, cell differentiation, cell signaling, and cell motility. Most growth and differentiation factors act on cells in their local environment (paracrine signaling). There are three broad classes of growth and differentiation factors. The first class includes the large polypeptide growth factors such as epidermal growth factor, fibroblast growth factor, transforming growth factor, insulin-like growth factor, and platelet-derived growth factor. The second class includes the hematopoietic growth factors such as the colony stimulating factors (CSFs). Hematopoietic growth factors stimulate the proliferation and differentiation of blood cells such as B-lymphocytes, T-lymphocytes, erythrocytes, platelets, eosinophils, basophils, neutrophils, macrophages, and their stem cell precursors. The third class includes small peptide factors such as bombesin, vasopressin, oxytocin, endothelin, transferrin, angiotensin II, vasoactive intestinal peptide, and bradykinin, which function as hormones to regulate cellular functions other than proliferation.

Growth and differentiation factors play critical roles in neoplastic transformation of cells *in vitro* and in tumor progression *in vivo*. Inappropriate expression of growth factors by tumor cells may contribute to vascularization and metastasis of tumors. During hematopoiesis, growth factor misregulation can result in anemias, leukemias, and lymphomas. Certain growth factors such as interferon are cytotoxic to tumor cells both *in vivo* and *in vitro*. Moreover, some growth factors and growth factor receptors are related both structurally and functionally to oncoproteins. In addition, growth factors affect transcriptional regulation of both proto-oncogenes and oncosuppressor genes. (Reviewed in Pimentel, E. (1994) Handbook of Growth Factors, CRC Press, Ann Arbor, MI, pp. 1-9.)

The Slit protein, first identified in Drosophila, is critical in central nervous system midline

formation and potentially in nervous tissue histogenesis and axonal pathfinding. Itoh et al. (1998; Brain Res. Mol. Brain Res. 62:175-186) have identified mammalian homologues of the slit gene (human Slit-1, Slit-2, Slit-3 and rat Slit-1). The encoded proteins are putative secreted proteins containing EGF-like motifs and leucine-rich repeats, both of which are conserved protein-protein interaction domains. Slit-1, -2, and -3 mRNAs are expressed in the brain, spinal cord, and thyroid, respectively (Itoh et al., *supra*). The Slit family of proteins are indicated to be functional ligands of glypican-1 in nervous tissue and it is suggested that their interactions may be critical in certain stages during central nervous system histogenesis (Liang, Y. et al. (1999) J. Biol. Chem. 274:17885-17892).

Neuropeptides and vasomediators (NP/VM) comprise a large family of endogenous signaling molecules. Included in this family are neuropeptides and neuropeptide hormones such as bombesin, neuropeptide Y, neurotensin, neuromedin N, melanocortins, opioids, galanin, somatostatin, tachykinins, urotensin II and related peptides involved in smooth muscle stimulation, vasopressin, vasoactive intestinal peptide, and circulatory system-borne signaling molecules such as angiotensin, complement, calcitonin, endothelins, formyl-methionyl peptides, glucagon, cholecystokinin and gastrin. NP/VMs can transduce signals directly, modulate the activity or release of other neurotransmitters and hormones, and act as catalytic enzymes in cascades. The effects of NP/VMs range from extremely brief to long-lasting. (Reviewed in Martin, C.R. et al. (1985) Endocrine Physiology, Oxford University Press, New York, NY, pp. 57-62.)

NP/VMs are involved in numerous neurological and cardiovascular disorders. For example, neuropeptide Y is involved in hypertension, congestive heart failure, affective disorders, and appetite regulation. Somatostatin inhibits secretion of growth hormone and prolactin in the anterior pituitary, as well as inhibiting secretion in intestine, pancreatic acinar cells, and pancreatic beta-cells. A reduction in somatostatin levels has been reported in Alzheimer's disease and Parkinson's disease. Vasopressin acts in the kidney to increase water and sodium absorption, and in higher concentrations stimulates contraction of vascular smooth muscle, platelet activation, and glycogen breakdown in the liver. Vasopressin and its analogues are used clinically to treat diabetes insipidus. Endothelin and angiotensin are involved in hypertension, and drugs, such as captopril, which reduce plasma levels of angiotensin, are used to reduce blood pressure (Watson, S. and S. Arkinstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 194; 252; 284; 55; 111).

Neuropeptides have also been shown to have roles in nociception (pain). Vasoactive intestinal peptide appears to play an important role in chronic neuropathic pain. Nociceptin, an endogenous ligand for for the opioid receptor-like 1 receptor, is thought to have a predominantly anti-nociceptive effect, and has been shown to have analgesic properties in different animal models of tonic or chronic pain (Dickinson, T. and S.M. Fleetwood-Walker (1998) Trends Pharmacol. Sci. 19:346-348).

Other proteins that contain signal peptides include secreted proteins with enzymatic activity. Such activity includes, for example, oxidoreductase/dehydrogenase activity, transferase activity, hydrolase activity, lyase activity, isomerase activity, or ligase activity. For example, matrix metalloproteinases are secreted hydrolytic enzymes that degrade the extracellular matrix and thus play an important role in tumor metastasis, tissue morphogenesis, and arthritis (Reponen, P. et al. (1995) Dev. Dyn. 202:388-396; Firestein, G.S. (1992) Curr. Opin. Rheumatol. 4:348-354; Ray, J.M. and W.G. Stetler-Stevenson (1994) Eur. Respir. J. 7:2062-2072; and Mignatti, P. and D.B. Rifkin (1993) Physiol. Rev. 73:161-195). Additional examples are the acetyl-CoA synthetases which activate acetate for use in lipid synthesis or energy generation (Luong, A. et al. (2000) J. Biol. Chem. 275:26458-26466). The result of acetyl-CoA synthetase activity is the formation of acetyl-CoA from acetate and CoA. Acetyl-CoA sythetases share a region of sequence similarity identified as the AMP-binding domain signature. Acetyl-CoA synthetase has been shown to be associated with hypertension (Toh, H. (1991) Protein Seq. Data Anal. 4:111-117; and Iwai, N. et al. (1994) Hypertension 23:375-380).

A number of isomerases catalyze steps in protein folding, phototransduction, and various anabolic and catabolic pathways. One class of isomerases is known as peptidyl-prolyl *cis-trans* isomerases (PPIases). PPIases catalyze the *cis* to *trans* isomerization of certain proline imidic bonds in proteins. Two families of PPIases are the FK506 binding proteins (FKBPs), and cyclophilins (CyPs). FKBPs bind the potent immunosuppressants FK506 and rapamycin, thereby inhibiting signaling pathways in T-cells. Specifically, the PPIase activity of FKBPs is inhibited by binding of FK506 or rapamycin. There are five members of the FKBP family which are named according to their calculated molecular masses (FKBP12, FKBP13, FKBP25, FKBP52, and FKBP65), and localized to different regions of the cell where they associate with different protein complexes (Coss, M. et al. (1995) J. Biol. Chem. 270:29336-29341; Schreiber, S.L. (1991) Science 251:283-287).

The peptidyl-prolyl isomerase activity of CyP may be part of the signaling pathway that leads to T-cell activation. CyP isomerase activity is associated with protein folding and protein trafficking, and may also be involved in assembly/disassembly of protein complexes and regulation of protein activity. For example, in *Drosophila*, the CyP NinaA is required for correct localization of rhodopsins, while a mammalian CyP (Cyp40) is part of the Hsp90/Hsc70 complex that binds steroid receptors. The mammalian CypA has been shown to bind the *gag* protein from human immunodeficiency virus 1 (HIV-1), an interaction that can be inhibited by cyclosporin. Since cyclosporin has potent anti-HIV-1 activity, CypA may play an essential function in HIV-1 replication. Finally, Cyp40 has been shown to bind and inactivate the transcription factor c-Myb, an effect that is reversed by cyclosporin. This effect implicates CyPs in the regulation of transcription, transformation, and differentiation (Bergsma, D.J. et al (1991) J. Biol. Chem. 266:23204-23214;

Hunter, T. (1998) Cell 92:141-143; and Leverson, J.D. and S.A. Ness, (1998) Mol. Cell. 1:203-211).

Gamma-carboxyglutamic acid (Gla) proteins rich in proline (PRGPs) are members of a family of vitamin K-dependent single-pass integral membrane proteins. These proteins are characterized by an extracellular amino terminal domain of approximately 45 amino acids rich in Gla. The intracellular carboxyl terminal region contains one or two copies of the sequence PPXY, a motif present in a variety of proteins involved in such diverse cellular functions as signal transduction, cell cycle progression, and protein turnover (Kulman, J.D. et al. (2001) Proc. Natl. Acad. Sci. USA 98:1370-1375). The process of post-translational modification of glutamic residues to form Gla is Vitamin K-dependent carboxylation. Proteins which contain Gla include plasma proteins involved in blood coagulation. These proteins are prothrombin, proteins C, S, and Z, and coagulation factors VII, IX, and X. Osteocalcin (bone-Gla protein, BGP) and matrix Gla-protein (MGP) also contain Gla (Friedman, P.A. and C.T. Przysiecki (1987) Int. J. Biochem. 19:1-7; Vermeer, C. (1990) Biochem. J. 266:625-636).

Immunoglobulins

Antigen recognition molecules are key players in the sophisticated and complex immune systems which all vertebrates have developed to provide protection from viral, bacterial, fungal, and parasitic infections. A key feature of the immune system is its ability to distinguish foreign molecules, or antigens, from "self" molecules. This ability is mediated primarily by secreted and transmembrane proteins expressed by leukocytes (white blood cells) such as lymphocytes, granulocytes, and monocytes. Most of these proteins belong to the immunoglobulin (Ig) superfamily, members of which contain one or more repeats of a conserved structural domain. This Ig domain is comprised of antiparallel β sheets joined by a disulfide bond in an arrangement called the Ig fold. The criteria for a protein to be a member of the Ig superfamily is to have one or more Ig domains, which are regions of 70-110 amino acid residues in length homologous to either Ig variable-like (V) or Ig constant-like (C) domains. Members of the Ig superfamily include antibodies (Ab), T cell receptors (TCRs), class I and II major histocompatibility (MHC) proteins and immune cell-specific surface markers such as the "cluster of differentiation" or CD antigens, CD2, CD3, CD4, CD8, poly-Ig receptors, Fc receptors, neural cell-adhesion molecule (NCAM) and platelet-derived growth factor receptor (PDGFR).

Ig domains (V and C) are regions of conserved amino acid residues that give a polypeptide a globular tertiary structure called an immunoglobulin (or antibody) fold, which consists of two approximately parallel layers of β-sheets. Conserved cysteine residues form an intrachain disulfide-bonded loop, 55-75 amino acid residues in length, which connects the two layers of β-sheets. Each β-sheet has three or four anti-parallel β-strands of 5-10 amino acid residues. Hydrophobic and hydrophilic interactions of amino acid residues within the β-strands stabilize the Ig fold (hydrophobic

on inward facing amino acid residues and hydrophilic on the amino acid residues in the outward facing portion of the strands). A V domain consists of a longer polypeptide than a C domain, with an additional pair of β-strands in the Ig fold.

A consistent feature of Ig superfamily genes is that each sequence of an Ig domain is encoded by a single exon. It is possible that the superfamily evolved from a gene coding for a single Ig domain involved in mediating cell-cell interactions. New members of the superfamily then arose by exon and gene duplications. Modern Ig superfamily proteins contain different numbers of V and/or C domains. Another evolutionary feature of this superfamily is the ability to undergo DNA rearrangements, a unique feature retained by the antigen receptor members of the family.

Many members of the Ig superfamily are integral plasma membrane proteins with extracellular Ig domains. The hydrophobic amino acid residues of their transmembrane domains and their cytoplasmic tails are very diverse, with little or no homology among Ig family members or to known signal-transducing structures. There are exceptions to this general superfamily description. For example, the cytoplasmic tail of PDGFR has tyrosine kinase activity. In addition Thy-1 is a glycoprotein found on thymocytes and T cells. This protein has no cytoplasmic tail, but is instead attached to the plasma membrane by a covalent glycophosphatidylinositol linkage.

Another common feature of many Ig superfamily proteins is the interactions between Ig domains which are essential for the function of these molecules. Interactions between Ig domains of a multimeric protein can be either homophilic or heterophilic (i.e., between the same or different Ig domains). Antibodies are multimeric proteins which have both homophilic and heterophilic interactions between Ig domains. Pairing of constant regions of heavy chains forms the Fc region of an antibody and pairing of variable regions of light and heavy chains form the antigen binding site of an antibody. Heterophilic interactions also occur between Ig domains of different molecules. These interactions provide adhesion between cells for significant cell-cell interactions in the immune system and in the developing and mature nervous system. (Reviewed in Abbas, A.K. et al. (1991) Cellular and Molecular Immunology, W.B. Saunders Company, Philadelphia, PA, pp. 142-145.)

Antibodies

MHC proteins are cell surface markers that bind to and present foreign antigens to T cells. MHC molecules are classified as either class I or class II. Class I MHC molecules (MHC I) are expressed on the surface of almost all cells and are involved in the presentation of antigen to cytotoxic T cells. For example, a cell infected with virus will degrade intracellular viral proteins and express the protein fragments bound to MHC I molecules on the cell surface. The MHC I/antigen complex is recognized by cytotoxic T-cells which destroy the infected cell and the virus within. Class II MHC molecules are expressed primarily on specialized antigen-presenting cells of the immune system, such as B-cells and macrophages. These cells ingest foreign proteins from the extracellular

fluid and express MHC II/antigen complex on the cell surface. This complex activates helper T-cells, which then secrete cytokines and other factors that stimulate the immune response. MHC molecules also play an important role in organ rejection following transplantation. Rejection occurs when the recipient's T-cells respond to foreign MHC molecules on the transplanted organ in the same way as to self MHC molecules bound to foreign antigen. (Reviewed in Alberts et al., *supra*, pp. 1229-1246.)

Antibodies are multimeric members of the Ig superfamily which are either expressed on the surface of B-cells or secreted by B-cells into the circulation. Antibodies bind and neutralize foreign antigens in the blood and other extracellular fluids. The prototypical antibody is a tetramer consisting of two identical heavy polypeptide chains (H-chains) and two identical light polypeptide chains (L-chains) interlinked by disulfide bonds. This arrangement confers the characteristic Y-shape to antibody molecules. Antibodies are classified based on their H-chain composition. The five antibody classes, IgA, IgD, IgE, IgG and IgM, are defined by the α, δ, ε, γ, and μ H-chain types. There are two types of L-chains, κ and λ, either of which may associate as a pair with any H-chain pair. IgG, the most common class of antibody found in the circulation, is tetrameric, while the other classes of antibodies are generally variants or multimers of this basic structure.

H-chains and L-chains each contain an N-terminal variable region and a C-terminal constant region. The constant region consists of about 110 amino acids in L-chains and about 330 or 440 amino acids in H-chains. The amino acid sequence of the constant region is nearly identical among H- or L-chains of a particular class. The variable region consists of about 110 amino acids in both H- and L-chains. However, the amino acid sequence of the variable region differs among H- or L-chains of a particular class. Within each H- or L-chain variable region are three hypervariable regions of extensive sequence diversity, each consisting of about 5 to 10 amino acids. In the antibody molecule, the H- and L-chain hypervariable regions come together to form the antigen recognition site. (Reviewed in Alberts et al. *supra*, pp. 1206-1213; 1216-1217.)

Both H-chains and L-chains contain the repeated Ig domains of members of the Ig superfamily. For example, a typical H-chain contains four Ig domains, three of which occur within the constant region and one of which occurs within the variable region and contributes to the formation of the antigen recognition site. Likewise, a typical L-chain contains two Ig domains, one of which occurs within the constant region and one of which occurs within the variable region.

The immune system is capable of recognizing and responding to any foreign molecule that enters the body. Therefore, the immune system must be armed with a full repertoire of antibodies against all potential antigens. Such antibody diversity is generated by somatic rearrangement of gene segments encoding variable and constant regions. These gene segments are joined together by site-specific recombination which occurs between highly conserved DNA sequences that flank each gene segment. Because there are hundreds of different gene segments, millions of unique genes can be

generated combinatorially. In addition, imprecise joining of these segments and an unusually high rate of somatic mutation within these segments further contribute to the generation of a diverse antibody population.

Parkinson's Disease

Parkinson's disease (PD) is a neurodegenerative disorder characterized by the progressive degeneration of the dopaminergic nigrostriatal pathway, and the presence of Lewy bodies. Genetic linkages for the Parkin gene to chromosome 6q25.2-27, for PARK3 to chromosome 2p (West, A. B. (2001) Eur. J. Hum. Genet. 9:659-666), and for PARK6 to chromosome 1p35-p36 have been identified (Valente, E. M. et al. (2002) Ann. Neurol. 51:14-18). Clinical disorders classified as parkinsonism include PD, dementia with Lewy bodies (DLB), progressive supranuclear palsy (PSP), and essential tremor. Several neurodegenerative diseases share pathogenic mechanisms involving tau or synuclein aggregation. These disorders include Alzheimer's disease, and Pick's disease as well as PD and progressive supranuclear palsy (Hardy, J. (2001) J. Alzheimers Dis. 3:109-116). Several genetically distinct forms of PD can be caused by mutations in single genes. Genes for monogenically inherited forms of Parkinson's disease have been mapped and/or cloned. In some families with autosomal dominant inheritance and typical Lewy-body pathology, mutations have been identified in the gene for alpha-synuclein. Aggregation of this protein in Lewy-bodies may be a crucial step in the molecular pathogenesis of familial and sporadic PD.

Parkin-mutations appear to be a common cause of PD in patients with very early onset. Mutations in the parkin gene of early-onset PD are autosomal recessive mutations in which nigral degeneration is not accompanied by Lewy-body formation. Parkin has been implicated in the cellular protein degradation pathways, as it has been shown that it functions as a ubiquitin ligase. A mutation in the gene for ubiquitin C-terminal hydrolase L1 in this pathway has been identified in another small family with PD. Other loci have been mapped to chromosome 2p and 4p, respectively, in families with dominantly inherited PD. These early-onset forms differ from the common sporadic form of PD. It is widely believed that a combination of interacting genetic and environmental causes may be responsible in the majority of PD-cases (Gasser, T. (2001) J. Neurol. 2001 248:833-840).

Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the

expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder. The potential application of gene expression profiling is particularly relevant to improving diagnosis, prognosis, and treatment of cancers, such as prostate, lung, ovarian, breast, bone, and colon cancers.

Prostate Cancer

Prostate cancer is a common malignancy in men over the age of 50, and the incidence increases with age. In the US, there are approximately 132,000 newly diagnosed cases of prostate cancer and more than 33,000 deaths from the disorder each year.

Once cancer cells arise in the prostate, they are stimulated by testosterone to a more rapid growth. Thus, removal of the testes can indirectly reduce both rapid growth and metastasis of the cancer. Over 95 percent of prostatic cancers are adenocarcinomas which originate in the prostatic acini. The remaining 5 percent are divided between squamous cell and transitional cell carcinomas, both of which arise in the prostatic ducts or other parts of the prostate gland.

As with most tumors, prostate cancer develops through a multistage progression ultimately resulting in an aggressive tumor phenotype. The initial step in tumor progression involves the hyperproliferation of normal luminal and/or basal epithelial cells. Androgen responsive cells become hyperplastic and evolve into early-stage tumors. Although early-stage tumors are often androgen sensitive and respond to androgen ablation, a population of androgen independent cells evolve from the hyperplastic population. These cells represent a more advanced form of prostate tumor that may become invasive and potentially become metastatic to the bone, brain, or lung. A variety of genes may be differentially expressed during tumor progression. For example, loss of heterozygosity (LOH) is frequently observed on chromosome 8p in prostate cancer. Fluorescence in situ hybridization (FISH) revealed a deletion for at least 1 locus on 8p in 29 (69%) tumors, with a significantly higher frequency of the deletion on 8p21.2-p21.1 in advanced prostate cancer than in localized prostate cancer, implying that deletions on 8p22-p21.3 play an important role in tumor differentiation, while 8p21.2-p21.1 deletion plays a role in progression of prostate cancer (Oba, K. et al. (2001) Cancer Genet. Cytogenet. 124: 20-26).

A primary diagnostic marker for prostate cancer is prostate specific antigen (PSA). PSA is a tissue-specific serine protease almost exclusively produced by prostatic epithelial cells. The quantity of PSA correlates with the number and volume of the prostatic epithelial cells, and consequently, the levels of PSA are an excellent indicator of abnormal prostate growth. Men with prostate cancer

exhibit an early linear increase in PSA levels followed by an exponential increase prior to diagnosis. However, since PSA levels are also influenced by factors such as inflammation, androgen and other growth factors, some scientists maintain that changes in PSA levels are not useful in detecting individual cases of prostate cancer.

Current areas of cancer research provide additional prospects for markers as well as potential therapeutic targets for prostate cancer. Several growth factors have been shown to play a critical role in tumor development, growth, and progression. The growth factors Epidermal Growth Factor (EGF), Fibroblast Growth Factor (FGF), and Tumor Growth Factor alpha (TGFα) are important in the growth of normal as well as hyperproliferative prostate epithelial cells, particularly at early stages of tumor development and progression, and affect signaling pathways in these cells in various ways (Lin J et al. (1999) Cancer Res. 59:2891-2897; Putz T et al. (1999) Cancer Res 59:227-233). The TGF-β family of growth factors are generally expressed at increased levels in human cancers and the high expression levels in many cases correlates with advanced stages of malignancy and poor survival (Gold LI (1999) Crit Rev Oncog 10:303-360). Finally, there are human cell lines representing both the androgen-dependent stage of prostate cancer (LNCap) as well as the androgen-independent, hormone refractory stage of the disease (PC3 and DU-145) that have proved useful in studying gene expression patterns associated with the progression of prostate cancer, and the effects of cell treatments on these expressed genes (Chung TD (1999) Prostate 15:199-207).

PrEC is a primary prostate epithelial cell line isolated from a normal donor. DU 145 is a prostate carcinoma cell line isolated from a metastatic site in the brain of 69-year old male with widespread metastatic prostate carcinoma. DU 145 has no detectable sensitivity to hormones; forms colonies in semi-solid medium; is only weakly positive for acid phosphatase; and cells are negative for prostate specific antigen (PSA).

LNCaP is a prostate carcinoma cell line isolated from a lymph node biopsy of a 50-year-old male with metastatic prostate carcinoma. LNCaP cells express prostate specific antigens, produce prostatic acid phosphatase, and express androgen receptors. PC-3 is a prostate adenocarcinoma cell line that was isolated from a metastatic site in the bone of a 62-year-old male with grade IV prostate adenocarcinoma.

Ovarian Cancer

Ovarian cancer is the leading cause of death from a gynecologic cancer. The majority of ovarian cancers are derived from epithelial cells, and 70% of patients with epithelial ovarian cancers present with late-stage disease. As a result, the long-term survival rates for this disease is very low. Identification of early-stage markers for ovarian cancer would significantly increase the survival rate. Genetic variations involved in ovarian cancer development include mutation of p53 and microsatellite instability. Gene expression patterns likely vary when normal ovary is compared to ovarian tumors.

Colon Cancer

Colon cancer evolves through a multi-step process whereby pre-malignant colonocytes undergo a relatively defined sequence of events leading to tumor formation. While soft tissue sarcomas are relatively rare, more than 50% of new patients diagnosed with the disease will die from it. The molecular pathways leading to the development of sarcomas are relatively unknown, due to the rarity of the disease and variation in pathology. Several factors participate in the process of tumor progression and malignant transformation including genetic factors, mutations, and selection.

In another pathway, it is believed that stem cells in the colonic crypts first undergo a primary hit, such as mutation of the APC gene. These mutated stem cells are then believed to grow abnormally and show defects in cell migration and adhesion. As a result, a polyp develops in the colon. The steps involved in the both the development of a polyp and the progression from a polyp to a colon tumor are not completely understood. There are likely many changes in gene expression that occur along this pathway, some of which may directly influence the development of cancer or are the result of cancer progression. In order to identify genes differentially expressed in colon cancer, we compared gene expression patterns in normal colon tissue, colon polyps, and colon tumors. Experiments have focused on determining the changes in gene expression that may influence the progression from normal colon tissue to polyp to cancer.

To understand the nature of gene alterations in colorectal cancer, a number of studies have focused on the inherited syndromes. Familial adenomatous polyposis (FAP), is caused by mutations in the adenomatous polyposis coli gene (APC), resulting in truncated or inactive forms of the protein. This tumor suppressor gene has been mapped to chromosome 5q. Hereditary nonpolyposis colorectal cancer (HNPCC) is caused by mutations in mis-match repair genes. Although hereditary colon cancer syndromes occur in a small percentage of the population and most colorectal cancers are considered sporadic, knowledge from studies of the hereditary syndromes can be generally applied. For instance, somatic mutations in APC occur in at least 80% of sporadic colon tumors. APC mutations are thought to be the initiating event in the disease. Other mutations occur subsequently. Approximately 50% of colorectal cancers contain activating mutations in ras, while 85% contain inactivating mutations in p53. Changes in all of these genes lead to gene expression changes in colon cancer.

PMA

The potential application of gene expression profiling is relevant to improving diagnosis, prognosis, and treatment of disease, such as diseases that affect immune responses. The immune system is made up of many cell types, that together orchestrate immune responses. Leukocytes, one class of immune cells, comprise lymphocytes, granulocytes, and monocytes. Lymphocytes include T- and B-cells, which specifically recognize and respond to foreign pathogens. T-cells fight viral infections and activate other leukocytes, while B-cells secrete antibodies that neutralize bacteria and

other microbes. Granulocytes and monocytes are primarily migratory, phagocytic cells that exit the bloodstream to fight infection in tissues. Monocytes, which are derived from immature promonocytes, further differentiate into macrophages that engulf and digest microorganisms and damaged or dead cells. For example, THP-1 is a human promonocyte cell line that acquires monocyte characteristics after treatment with both phorbol ester such as phorbol myristate acetate (PMA). PMA is a broad activator of the protein kinase C-dependent pathways.

Monocytes and macrophages modulate the immune response by secreting signaling molecules such as growth factors and cytokines. Monocytes are involved in the initiation and maintenance of inflammatory immune responses. The outer membrane of gram-negative bacteria expresses lipopolysaccharide (LPS) complexes called endotoxins. Toxicity is associated with the lipid component (Lipid A) of LPS, and immunogenicity is associated with the polysaccharide components of LPS. LPS elicits a variety of inflammatory responses, and because it activates complement by the alternative (properdin) pathway, it is often part of the pathology of gram-negative bacterial infections. In addition, monocytes and macrophages are recruited to sites of infection and inflammation by signaling proteins secreted by other leukocytes.

Glucocorticoids

The C3A cell line is a clonal derivative of the Hep G2 hepatoma cell line isolated from a 15-year-old male with liver tumor, which was selected for its strong contact inhibition of growth. The use of a clonal population enhances the reproducibility of the cells. By maintaining a regularly monitored Master Bank and limited passages of cells, the robustness of the cell line for long-term study is greatly improved. The C3A cells express insulin receptor and insulin-like growth factor II receptor. Betamethasone [378-44-9] is a synthetic glucocorticoid used as anti-inflammatory or immunosuppressive agents. Glucocorticoids are naturally occurring hormones that prevent or suppress inflammation and immune responses when administered at pharmacological doses. At the molecular level, unbound glucocorticoids readily cross cell membranes and bind with high affinity to specific cytoplasmic receptors. Subsequent to binding, transcription and, ultimately, protein synthesis are affected. The result can include inhibition of leukocyte infiltration at the site of inflammation, interference in the function of mediators of inflammatory response, and suppression of humoral immune responses. The anti-inflammatory actions of corticosteroids are thought to involve phospholipase A2 inhibitory proteins, collectively called lipocortins. Lipocortins, in turn, control the biosynthesis of potent mediators of inflammation such as prostaglandins and leukotrienes by inhibiting the release of the precursor molecule arachidonic.

Tangier Disease

Tangier disease (TD) is a genetic disorder characterized by near absence of circulating high density lipoprotein (HDL) and the accumulation of cholesterol esters in many tissues, including

tonsils, lymph nodes, liver, spleen, thymus, and intestine. Low levels of HDL represent a clear predictor of premature coronary artery disease and homozygous TD correlates with a four- to six-fold increase in cardiovascular disease compared to controls. HDL plays a cardio-protective role in reverse cholesterol transport, the flux of cholesterol from peripheral cells such as tissue macrophages through plasma lipoproteins to the liver. The HDL protein, apolipoprotein A-I, plays a major role in this process, interacting with the cell surface to remove excess cholesterol and phospholipids. This pathway is severely impaired in TD and the defect lies in a specific gene, the ABC1 transporter. This gene is a member of the family of ATP-binding cassette transporters, which utilize ATP hydrolysis to transport a variety of substrates across membranes.

Breast Cancer

Breast cancer is the most frequently diagnosed type of cancer in American women and the second most frequent cause of cancer death. The lifetime risk of an American woman developing breast cancer is 1 in 8, and one-third of women diagnosed with breast cancer die of the disease. A number of risk factors have been identified, including hormonal and genetic factors. Many studies have focused on identifying the genetic abnormalities that occur in breast cancer cells. One genetic defect results in a loss of heterozygosity (LOH) at multiple loci. Some of the genes identified from these studies include p53, Rb, BRCA1, and BRCA2. Another genetic defect is gene amplification involving genes such as c-myc and c-erbB2 (Her2-neu gene. Steroid and growth factor pathways are also altered in breast cancer, notably the estrogen, progesterone, and epidermal growth factor (EGF) pathways. Histological and molecular evaluation of breast tumors have revealed that the development of breast cancer evolves through a multi-step process whereby pre-malignant mammary epithelial cells undergo a relatively defined sequence of events leading to tumor formation. An early event in tumor development is ductal hyperplasia. Cells undergoing rapid neoplastic growth gradually progress to invasive carcinoma and become metastatic to the lung, bone, and potentially other organs. Several factors participate in the process of tumor progression and malignant transformation, including genetic factors, environmental factors, growth factors, and hormones.

Osteosarcoma

Osteosarcoma is the most common malignant bone tumor in children. Approximately 80% of patients present with non-metastatic disease. After the diagnosis is made by an initial biopsy, treatment involves the use of 3–4 courses of neoadjuvant chemotherapy before definitive surgery, followed by post-operative chemotherapy. With currently available treatment regimens, approximately 30–40% of patients with non-metastatic disease relapse after therapy. Currently, there is no prognostic factor that can be used at the time of initial diagnosis to predict which patients will have a high risk of relapse. The only significant prognostic factor predicting the outcome in a patient with non-metastatic osteosarcoma is the histopathologic response of the primary tumor resected at the

time of definitive surgery. The degree of necrosis in the primary tumor is a reflection of the tumor response to neoadjuvant chemotherapy. A higher degree of necrosis (good or favorable response) is associated with a lower risk of relapse and a better outcome. Patients with a lower degree of necrosis (poor or unfavorable response) have a much higher risk of relapse and poor outcome even after complete resection of the primary tumor. Unfortunately, poor outcome cannot be altered despite modification of post-operative chemotherapy to account for the resistance of the primary tumor to neoadjuvant chemotherapy. Thus, there is an urgent need to identify prognostic factors that can be used at the time of diagnosis to recognize the subtypes of osteosarcomas that have various risks of relapse, so that more appropriate chemotherapy can be used at the outset to improve the outcome.

PBMCs

The potential application of gene expression profiling is particularly relevant to improving the diagnosis, prognosis, and treatment of disease, such as diseases that affect immune responses. Human peripheral blood mononuclear cells (PBMCs) can be classified into discrete cellular populations representing the major cellular components of the immune system. PBMCs contain about
52% lymphocytes (12% B lymphocytes, 40% T lymphocytes {25% CD4+ and 15% CD8+}), 20% NK cells, 25% monocytes, and 3% various cells that include dendritic cells and progenitor cells. The proportions, as well as the biology of these cellular components tend to vary slightly between healthy individuals, depending on factors such as age, gender, past medical history, and genetic background. Staphylococcal exotoxins specifically activate human T cells, expressing an appropriate TCR-Vbeta chain. Although polyclonal in nature, T cells activated by Staphylococcal exotoxins require antigen presenting cells (APCs) to present the exotoxin molecules to the T cells and deliver the costimulatory signals required for optimum T cell activation. Although, Staphylococcal exotoxins must be presented to T cells by APCs, these molecules are not required to be processed by APC. Indeed, Staphylococcal exotoxins directly bind to a non-polymorphic portion of the human MHC class II molecules, bypassing the need for capture, cleavage, and binding of the peptides to the polymorphic antigenic groove of the MHC class II molecules. PMA is a broad activator of the protein kinase C-dependent pathways. Ionomycin is a calcium ionophore that permits the entry of calcium in the cell, hence increasing the cytosolic calcium concentration. The combination of PMA and ionomycin activates two of the major signaling pathways used by mammalian cells to interact with their environment. In PBMCs, the combination of PMA and ionomycin mimics the secondary signaling events elicited during activation of lymphocytes, NK cells, and monocytes.

Interleukins

Interleukin 1 beta (IL-1b) is a cytokine associated with acute inflammatory responses and is generally considered the prototypical pro-inflammatory cytokine. However, IL-1 β functions are not

limited to the inflammatory response since this molecule is involved in processes such as fever induction, metabolic regulation, and bone remodeling. Both cells of the immune system (monocytes, dendritic cells, NK cells, platelets, and neutrophils) and somatic cells (osteoblasts, neurons, Schwann's cells, oligodendrocytes, and adrenal cortical cells) can produce IL-1 β. IL-1 β has been shown to induce its own production in monocytes; induce the production of adhesion molecules and chemokines in endothelial cells; and in conjunction with IL-12, induce interferon-γ production by NK Cells. IL-1 β is produced as a single chain pro-molecule that needs to be cleaved by a specialized protease, IL-1 β Converting Enzyme (ICE), to acquire its function.

Interleukin 2 (IL-2) is a protein with a variety of immunologic functions, most notably the ability to promote the proliferation and maturation of activated T cells. Some of the biological activities attributed to IL-2 include: induction of secretion of IFN-γ and TNF-α and -β from PBMCs; stimulation of the rate of synthesis of c-myc RNA and transferrin receptor; activation of neutrophils; stimulation of proliferation and maturation of activated helper T cells; stimulation of proliferation of activated and natural killer cells and tumor-infiltrating lymphocytes, as well as enhancement of the ability to kill target cells; induction of IL-2 receptor expression on T cells; stimulation of antibody-producing B cell proliferation.

Interleukin 3 (IL-3) is a pleiotropic factor produced primarily by activated T cells that can stimulate the proliferation and differentiation of pluripotent hematopoietic stem cells and various lineage committed progenitors. IL-3 also affects the functional activity of mature mast cells, basophils, eosinophils, and macrophages. Because of its multiple functions and targets, IL-3 was originally studied under different names, including mast cell growth factor, P-cell stimulating factor, burst promoting activity, multi-colony stimulating factor, thy-1 inducing factor, and WEHI-3 growth factor. In addition to activated T cells, other cell types such as human thymic epithelial cells, activated murine mast cells, murine keratinocytes, and neurons/astrocytes can also produce IL-3. IL-3 exerts its biological activities by binding to specific cell surface receptors. The high affinity receptor responsible for IL-3 signaling is composed of at least two subunits, an IL-3 specific α-chain that binds IL-3 with low affinity and a common β-chain that is shared by the IL-5 and GM-CSF high affinity receptors. Although the β-chain itself does not bind IL-3, it confers high-affinity IL-3 binding in the presence of the α-chain. Receptors for IL-3 are present on bone marrow progenitors, macrophages, mast cells, eosinophils, megakaryocytes, basophils, and various myeloid leukemic cells.

Interleukin 4 (IL-4) is a pleiotropic cytokine produced by activated T cells, mast cells, and basophils. It was initially identified as a B cell differentiation factor (BCDF) and a B cell stimulatory factor (BSF1). Subsequent to the molecular cloning and expression of both human and mouse IL-4, numerous other functions have been ascribed to B cells and other hematopoietic and non-hematopoietic cells including T lymphocytes, monocytes, macrophages, mast cells, myeloid and

erythroid progenitors, fibroblasts, endothelial cells, etc. IL-4 exhibits anti-tumor effects both *in vivo* and *in vitro*. Recently, IL-4 was identified as an important regulator for the CD4+ subset (Th1-like vs. Th2-like) development. The biological effects of IL-4 are mediated by the binding of IL-4 to specific cell surface receptors. The functional high-affinity receptor for IL-4 consists of a ligand-binding subunit (IL-4 R) and a second subunit (β chain) that can modulate the ligand binding affinity of the receptor complex. In certain cell types, the gamma chain of the IL-2 receptor complex is a functional β chain of the IL-4 receptor complex. Signaling of IL-4 through its receptor leads to the activation of Signal Transducer and Activator of Transcription 6 (STAT6).

Interleukin 5 (IL-5) is a T cell-derived factor that promotes the proliferation, differentiation, and activation of eosinophils. IL-5 has also been known as T cell replacing factor (TRF), B cell growth factor II (BCGFII), B cell differentiation factor m (BCDF m), eosinophil differentiation factor (EDF), and eosinophil colony-stimulating factor (Eo-CSF). IL-5 exerts its activity on target cells by binding to specific cell surface receptors. The functional high-affinity receptor for human IL-5 is composed of a low-affinity IL-5 binding α-subunit and a non-binding common β-subunit that is shared with the high-affinity receptors for GM-CSF and IL-3.

Interleukin 6 (IL-6) is a multifunctional protein that plays important roles in host defense, acute phase reactions, immune responses, and hematopoiesis. According to the type of biological responses being studied, IL-6 was previously named interferon-b2, 26-kDa protein, B cell stimulatory factor-2 (BSF-2), hybridoma/plasmacytoma growth factor, hepatocyte stimulating factor, cytotoxic T cell differentiation factor, and macrophage-granulocyte inducing factor 2A (MGI-2A). The IL-6 designation was adopted after these variously named proteins were found to be identical on the basis of their amino acid and/or nucleotide sequences. IL-6 is expressed by a variety of normal and transformed cells including T cells, B cells, monocytes/macrophages, fibroblasts, hepatocytes, keratinocytes, astrocytes, vascular endothelial cells, and various tumor cells. The production of IL-6 is upregulated by numerous signals including mitogenic or antigenic stimulation, LPS, calcium ionophore, IL-1, IL-2, IFN, TNF, PDGF, and viruses. IL-4 and IL-13 inhibit IL-6 expression in monocytes.

Interleukin 7 (IL-7), previously known as pre-B-cell growth factor and lymphopoietin-1, was originally purified on the basis of its ability to promote the proliferation of precursor B-cells. It has been shown that IL-7 can also stimulate the proliferation of thymocytes, T cell progenitors, and mature CD4 + and CD8 + T cells. IL-7 can induce the formation of lymphokine-activated killer (LAK) cells as well as the development of cytotoxic T lymphocytes (CTL). Among myeloid lineage cells, IL-7 can upregulate the production of pro-inflammatory cytokines and stimulate the tumoricidal activity of monocytes/ macrophages. IL-7 is expressed by adherent stromal cells from various tissues. IL-7 bioactivities are mediated by the binding of IL-7 to functional high-affinity receptor complexes.

The ligand binding subunit (IL-7 R) of the IL-7 receptor complex has been cloned from human and mouse sources. Recently, the γ chain of the IL-2 receptor complex has been shown to be an essential component for IL-7 signal transduction. Both IL-7 R and IL-2 R γ are members of the hematopoietin receptor superfamily. Cells known to express IL-7 receptors include pre-B cells, T cells, and bone marrow cells.

Interleukin 8 (IL-8) was originally discovered and purified independently by a number of laboratories as a neutrophil chemotactic and activating factor. It was also referred to as neutrophil chemotactic factor (NCF), neutrophil activating protein (NAP), monocyte-derived neutrophil chemotactic factor (MDNCF), T-lymphocyte chemotactic factor (TCF), granulocyte chemotactic protein (GCP), and leukocyte adhesion inhibitor (LAI). Many cell types, including monocyte/macrophages, T cells, neutrophils, fibroblasts, endothelial cells, keratinocytes, hepatocytes, chondrocytes, and various tumor cell lines can produce IL-8 in response to a wide variety of pro-inflammatory stimuli such as exposure to IL-1, TNF, LPS, and viruses. IL-8 is a member of the alpha (C-X-C) subfamily of chemokines, which also includes platelet factor 4, GRO, IP-10, etc. IL-8 is a potent chemoattractant for neutrophils and has a wide range of other pro-inflammatory effects. IL-8 causes degranulation of neutrophil-specific granules and azurophilic granules. IL-8 induces expression of the cell adhesion molecules CD11/ CD18 and enhances the adherence of neutrophils to endothelial cells and sub-endothelial matrix proteins. Besides neutrophils, IL-8 is also chemotactic for basophils, T cells, and eosinophils. IL-8 has been reported to be a co-mitogen for keratinocytes and was also shown to be an autocrine growth factor for melanoma cells. Recently, IL-8 was reported to be angiogenic both *in vivo* and *in vitro*.

Interleukin 10 (IL-10), initially designated cytokine synthesis inhibitory factor (CSIF), was originally identified as a product of murine T helper 2 (Th2) clones that inhibited the cytokine production by Th1 clones, which are dependent upon stimulation with antigen in the presence of antigen presenting cells (APC). The human homolog of murine IL-10 was subsequently cloned by cross-hybridization. Human IL-10 is produced by CD4 + T cell clones as well as by some CD8 + T cell clones. In addition, human B cells, EBV-transformed lymphoblastoid cell lines, and monocytes can also produce IL-10 upon activation. IL-10 is a pleiotrophic cytokine that can exert either immunostimulatory or immunosupressive effects on a variety of cell types. It is a potent immunosuppressant of macrophage functions. *In vitro*, IL-10 can inhibit the accessory function and antigen-presenting capacity of monocytes by, among other effects, downregulating class II MHC expression. Thus, IL-10 can inhibit monocyte/ macrophage-dependent, antigen-specific proliferation of mouse Th1 clones as well as human Th0-, Th1-, and Th2-like T cells. IL-10 can also inhibit the monocyte/macrophage-dependent, antigen stimulated cytokine synthesis (especially IFN-γ) by human PBMNC and NK cells. Additionally, IL-10 is a potent inhibitor of monocyte/macrophage activation

and its resultant cytotoxic effects. It can suppress the production of numerous cytokines including TNF-α , IL-1, IL-6, and IL-10, as well as the synthesis of superoxide anion, reactive oxygen intermediates, and reactive nitrogen intermediates by activated monocytes/macrophages. As an immunostimulatory cytokine, IL-10 can act on B cells to enhance their viability, cell proliferation, Ig secretion, and class II MHC expression. Aside from B lymphocytes, IL-10 is also a growth co-stimulator for thymocytes and mast cells, as well as an enhancer of cytotoxic T cell development.

Interleukin 12 (IL-12), also known as natural killer cell stimulatory factor (NKSF) or cytotoxic lymphocyte maturation factor (CLMF), is a pleiotropic cytokine originally identified in the medium of activated human B lymphoblastoid cell lines. IL-12 is produced by macrophages and B lymphocytes and has been shown to have multiple effects on T cells and natural killer (NK) cells. These include inducing production of IFN-γ and TNF by resting and activated T and NK cells, enhancing the cytotoxic activity of resting NK and T cells, inducing and synergizing with IL-2 in the generation of lymphokine-activated killer (LAK) cells, acting as a comitogen to stimulate proliferation of resting T cells, and inducing proliferation of activated T and NK cells. Current evidence indicates that IL-12, produced by macrophages in response to infectious agents, is a central mediator of the cell-mediated immune response by its actions on the development, proliferation, and activities of TH1 cells. In its role as the initiator of cell-mediated immunity, it has been suggested that IL-12 has therapeutic potential as a stimulator of cell-mediated immune responses to microbial pathogens, metastatic cancers, and viral infections such as AIDS.

Interleukin 18 (IL-18), also known as interferon-gamma-inducing factor (IGIF) and IL-1 γ, is a recently described cytokine that shares some biologic activities with IL-12 and structural similarities with the IL-1 family of proteins. IL-18 was originally cloned from liver cells and has since been shown to be expressed by monocyte/macrophages, osteoblasts, and keratinocytes. Human IL-18 cDNA encodes a 193 amino acid residue biologically inactive precursor molecule (pro-IL-18) that requires cleavage by a specific protease -- ICE -- to acquire its function. Like IL-12, human IL-18 has been shown to enhance NK cell activity in PBMC cultures. Human IL-18 has also been found to induce the production of IFN-γ and GM-CSF while inhibiting the production of IL-10 by PBMCs. On enriched human T cells, human IL-18 can enhance Th1 cytokine production and stimulate cell proliferation via an IL-2-dependent pathway.

Granulocyte Colony Stimulating Factor (G-CSF) is a pleiotropic cytokine best known for its specific effects on the proliferation, differentiation, and activation of hematopoietic cells of the neutrophilic granulocyte lineage. Activated monocytes and macrophages are the primary sources of G-CSF in the body. Fibroblasts, endothelial cells, astrocytes, and bone marrow stromal cells can also produce this cytokine upon activation. *In vitro,* G-CSF stimulates growth, differentiation, and functions of cells from the neutrophil lineage. Consistent with its in vitro functions, G-CSF plays

important roles in defending against infection, in inflammation and repair, and in maintaining steady state hematopoiesis.

Granulocyte-Monocyte Colony Stimulating Factor (GM-CSF) Granulocyte-monocyte colony stimulating factor (GM-CSF) was first described as a factor that can support the *in vitro* colony formation of granulocyte-macrophage progenitors. In addition, GM-CSF is a growth factor for erythroid, megakaryocyte, and eosinophil progenitors. Lymphocytes (T and B), monocytes, macrophages, mast cells, endothelial cells, and fibroblasts can produce GM-CSF upon activation. GM-CSF exerts its biological effects by binding to specific cell surface receptors. The high affinity receptors required for human GM-CSF signal transduction are heterodimers consisting of a GM-CSF-specific α chain and a common β chain that is shared by the high-affinity receptors for IL-3 and IL-5.

Interferon gamma (IFN-γ), also known as Type II interferon or immune interferon, is a cytokine produced primarily by T-lymphocytes and natural killer cells. IFN-γ was originally characterized based on its antiviral activities. The protein also exerts antiproliferative, immunoregulatory, and proinflammatory activities and is thus important in host defense mechanisms. IFN-γ induces the production of cytokines; and upregulates the expression of class I and II MHC antigens, Fc receptor, and leukocyte adhesion molecules. It modulates macrophage effector functions, influences isotype switching, and potentiates the secretion of immunoglobulins by B cells. IFN-γ also augments TH1 cell expansion and may be required for TH1 cell differentiation. The IFN-γ receptor is present on almost all cell types except mature erythrocytes and has been cloned and characterized. The IFN-γ receptor is structurally related to the recently cloned IL-10 receptor.

Leptin is a protein product of the mouse obesity gene. Mice with mutations in the obesity gene that block the synthesis of leptin tend to be obese and diabetic and exhibit reduced activity, metabolism, and body temperature. Human leptin shares approximately 84% sequence identity with the mouse protein. Human leptin cDNA encodes a 167 amino acid residue protein with a 21 amino acid residue signal sequence that is cleaved to yield the 146 amino acid residue mature protein. The expression of leptin mRNA is restricted to adipose tissue. A high-affinity receptor for leptin (OB-R) with homology to gp130 and the G-CSF receptor has recently been cloned. OB-R mRNA is expressed in the choroid plexus and in the hypothalamus. OB-R is also an isoform of B219, a sequence that is expressed in at least four isoforms in very primitive hematopoietic cell populations and in a variety of lymphohematopoietic cell lines. The possible roles of leptin in body weight regulation, hematopoiesis, and reproduction are being investigated.

Leukemia inhibitory factor (LIF) was initially identified as a factor that inhibits the proliferation and induces the differentiation to macrophages of the murine myeloid leukemic cell line M1. Subsequent to its purification and molecular cloning, LIF was recognized to be a pleiotropic

factor with multiple effects on both hematopoietic and non-hematopoietic cells. LIF has overlapping biological functions with OSM, IL-6, IL-11, and CNTF. All these cytokines use gp130 as a component in their signal transducing receptor complexes. Human LIF cDNA encodes a 202 amino acid residue polypeptide with a 22 amino acid residue signal peptide that is cleaved to yield a 180 amino acid residue mature human LIF.

Tumor Growth Factor beta (TGF-β) is a stable, multifunctional polypeptide growth factor. While specific receptors for this protein have been found on almost all mammalian cell types thus far examined, the effect of the molecule varies depending on the cell type and growth conditions. Generally, TGF-β is stimulatory for cells of mesenchymal origin and inhibitory for cells of epithelial or neuroectodermal origin. TGF-β has been found in the highest concentration in human platelets and mammalian bone, but is produced by many cell types in smaller amounts.

Tumor necrosis factor alpha (TNF-α), also called cachectin, is produced by neutrophils, activated lymphocytes, macrophages, NK cells, LAK cells, astrocytes, endothelial cells, smooth muscle cells, and some transformed cells. TNF-α occurs as a secreted, soluble form and as a membrane-anchored form, both of which are biologically active. Two types of receptors for TNF-α have been described and virtually all cell types studied show the presence of one or both of these receptor types. TNF-α and TNF-β are extremely pleiotropic factors due to the ubiquity of their receptors, to their ability to activate multiple signal transduction pathways, and to their ability to induce or suppress the expression of a wide number of genes. TNF-α and TNF-β play a critical role in mediation of the inflammatory response and in mediation of resistance to infections and tumor growth.

Lung Cancer

Lung cancer is the leading cause of cancer death in the United States, affecting more than 100,000 men and 50,000 women each year. Nearly 90% of the patients diagnosed with lung cancer are cigarette smokers. Tobacco smoke contains thousands of noxious substances that induce carcinogen metabolizing enzymes and covalent DNA adduct formation in the exposed bronchial epithelium. In nearly 80% of patients diagnosed with lung cancer, metastasis has already occurred. Most commonly lung cancers metastasize to pleura, brain, bone, pericardium, and liver. The decision to treat with surgery, radiation therapy, or chemotherapy is made on the basis of tumor histology, response to growth factors or hormones, and sensitivity to inhibitors or drugs. With current treatments, most patients die within one year of diagnosis. Earlier diagnosis and a systematic approach to identification, staging, and treatment of lung cancer could positively affect patient outcome.

Lung cancers progress through a series of morphologically distinct stages from hyperplasia

to invasive carcinoma. Malignant lung cancers are divided into two groups comprising four histopathological classes. The Non Small Cell Lung Carcinoma (NSCLC) group includes squamous cell carcinomas, adenocarcinomas, and large cell carcinomas and accounts for about 70% of all lung cancer cases. Adenocarcinomas typically arise in the peripheral airways and often form mucin secreting glands. Squamous cell carcinomas typically arise in proximal airways. The histogenesis of squamous cell carcinomas may be related to chronic inflammation and injury to the bronchial epithelium, leading to squamous metaplasia. The Small Cell Lung Carcinoma (SCLC) group accounts for about 20% of lung cancer cases. SCLCs typically arise in proximal airways and exhibit a number of paraneoplastic syndromes including inappropriate production of adrenocorticotropin and anti-diuretic hormone. Lung cancer cells accumulate numerous genetic lesions, many of which are associated with cytologically visible chromosomal aberrations. The high frequency of chromosomal deletions associated with lung cancer may reflect the role of multiple tumor suppressor loci in the etiology of this disease. Deletion of the short arm of chromosome 3 is found in over 90% of cases and represents one of the earliest genetic lesions leading to lung cancer. Deletions at chromosome arms 9p and 17p are also common. Other frequently observed genetic lesions include overexpression of telomerase, activation of oncogenes such as K-ras and c-myc, and inactivation of tumor suppressor genes such as RB, p53 and CDKN2.

Genes differentially regulated in lung cancer have been identified by a variety of methods. Using mRNA differential display technology, Manda et al. (1999; Genomics 51:5-14) identified five genes differentially expressed in lung cancer cell lines compared to normal bronchial epithelial cells. Among the known genes, pulmonary surfactant apoprotein A and alpha 2 macroglobulin were down regulated whereas nm23H1 was upregulated. Petersen et al. (2000; Int J. Cancer, 86:512-517) used suppression subtractive hybridization to identify 552 clones differentially expressed in lung tumor derived cell lines, 205 of which represented known genes. Among the known genes, thrombospondin-1, fibronectin, intercellular adhesion molecule 1, and cytokeratins 6 and 18 were previously observed to be differentially expressed in lung cancers. Wang et al. (2000; Oncogene 19:1519-1528) used a combination of microarray analysis and subtractive hybridization to identify 17 genes differentially overexpresssed in squamous cell carcinoma compared with normal lung epithelium. Among the known genes they identified were keratin isoform 6, KOC, SPRC, IGFb2, connexin 26, plakofillin 1 and cytokeratin 13.

The potential application of gene expression profiling is particularly relevant to improving the diagnosis, prognosis, and treatment of diseases, such as those diseases that affect metabolism.

Adipocyte Metabolism

The potential application of gene expression profiling is particularly relevant to improving

diagnosis, prognosis, and treatment of disease, such as diseases that affect metabolism. The primary function of adipose tissue is the ability to store and release fat during periods of feeding and fasting. White adipose tissue is the major energy reserve in periods of fasting, and its reserve is mobilized during energy deprivation. Adipose tissue is one of the primary target tissues for insulin, and adipogenesis and insulin resistance are linked in type II diabetes, non-insulin dependent diabetes mellitus (NIDDM). Cytologically the conversion of a preadipocytes into mature adipocytes is characterized by deposition of fat droplets around the nuclei. The conversion process *in vivo* can be induced by thiazolidinediones (TZDs) and other PPARγ agonists (Adams et al. (1997) J. Clin. Invest. 100:3149-3153) which also lead to increased sensitivity to insulin and reduced plasma glucose and blood pressure.

Thiazolidinediones (TZDs) act as agonists for the peroxisome-proliferator-activated receptor gamma (PPARγ), a member of the nuclear hormone receptor superfamily. TZDs reduce hyperglycemia, hyperinsulinemia, and hypertension, in part by promoting glucose metabolism and inhibiting gluconeogenesis. Roles for PPARγ and its agonists have been demonstrated in a wide range of pathological conditions including diabetes, obesity, hypertension, atherosclerosis, polycystic ovarian syndrome, and cancers such as breast, prostate, liposarcoma, and colon cancer.

The mechanism by which TZDs and other PPARγ agonists enhance insulin sensitivity is not fully understood, but may involve the ability of PPARγ to promote adipogenesis. When ectopically expressed in cultured preadipocytes, PPARγ is a potent inducer of adipocyte differentiation. TZDs, in combination with insulin and other factors, can also enhance differentiation of human preadipocytes in culture (Adams et al. (1997) J. Clin. Invest. 100:3149-3153). The relative potency of different TZDs in promoting adipogenesis in vitro is proportional to both their insulin sensitizing effects *in vivo*, and their ability to bind and activate PPARγ *in vitro*. Interestingly, adipocytes derived from omental adipose depots are refractory to the effects of TZDs. It has therefore been suggested that the insulin sensitizing effects of TZDs may result from their ability to promote adipogenesis in subcutaneous adipose depots (Adams et al., *ibid*). Further, dominant negative mutations in the PPARγ gene have been identified in two non-obese subjects with severe insulin resistance, hypertension, and overt non-insulin dependent diabetes mellitus (NIDDM) (Barroso et al. (1998) Nature 402:880-883).

NIDDM is the most common form of diabetes mellitus, a chronic metabolic disease that affects 143 million people worldwide. NIDDM is characterized by abnormal glucose and lipid metabolism that result from a combination of peripheral insulin resistance and defective insulin secretion. NIDDM has a complex, progressive etiology and a high degree of heritability. Numerous complications of diabetes including heart disease, stroke, renal failure, retinopathy, and peripheral

neuropathy contribute to the high rate of morbidity and mortality.

At the molecular level, PPARγ functions as a ligand activated transcription factor. In the presence of ligand, PPARγ forms a heterodimer with the retinoid X receptor (RXR) which then activates transcription of target genes containing one or more copies of a PPARγ response element (PPRE). Many genes important in lipid storage and metabolism contain PPREs and have been identified as PPARγ targets, including PEPCK, aP2, LPL, ACS, and FAT-P (Auwerx, J. (1999) Diabetologia 42:1033-1049). Multiple ligands for PPARγ have been identified. These include a variety of fatty acid metabolites; synthetic drugs belonging to the TZD class, such as Pioglitazone and Rosiglitazone (BRL49653); and certain non-glitazone tyrosine analogs such as GI262570 and GW1929. The prostaglandin derivative 15-dPGJ2 is a potent endogenous ligand for PPARγ.

Expression of PPARγ is very high in adipose but barely detectable in skeletal muscle, the primary site for insulin stimulated glucose disposal in the body. PPARγ is also moderately expressed in large intestine, kidney, liver, vascular smooth muscle, hematopoietic cells, and macrophages. The high expression of PPARγ in adipose suggests that the insulin sensitizing effects of TZDs may result from alterations in the expression of one or more PPARγ regulated genes in adipose tissue. Identification of PPARγ target genes will contribute to better drug design and the development of novel therapeutic strategies for diabetes, obesity, and other conditions.

Systematic attempts to identify PPARγ target genes have been made in several rodent models of obesity and diabetes (Suzuki et al. (2000) Jpn. J. Pharmacol. 84:113-123; Way et al. (2001) Endocrinology 142:1269-1277). However, a serious drawback of the rodent gene expression studies is that significant differences exist between human and rodent models of adipogenesis, diabetes, and obesity (Taylor (1999) Cell 97:9-12; Gregoire et al. (1998) Physiol. Reviews 78:783-809). Therefore, an unbiased approach to identifying TZD regulated genes in primary cultures of human tissues is necessary to fully elucidate the molecular basis for diseases associated with PPARγ activity.

The majority of research in adipocyte biology to date has been done using transformed mouse preadipocyte cell lines. The culture condition, which stimulates mouse preadipocyte differentiation is different from that for inducing human primary preadipocyte differentiation. In addition, primary cells are diploid and may therefore reflect the *in vivo* context better than aneuploid cell lines. Understanding the gene expression profile during adipogenesis in human will lead to understanding the fundamental mechanism of adiposity regulation. Furthermore, through comparing the gene expression profiles of adipogenesis between donor with normal weight and donor with obesity, identification of crucial genes, potential drug targets for obesity and type II diabetes, will be possible.

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, cardiovascular,

neurological, developmental, and metabolic disorders.

SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, secreted proteins, referred to collectively as 'SECP' and individually as 'SECP-1,' 'SECP-2,' 'SECP-3,' 'SECP-4,' 'SECP-5,' 'SECP-6,' 'SECP-7,' 'SECP-8,' 'SECP-9,' 'SECP-10,' 'SECP-11,' 'SECP-12,' 'SECP-13,' 'SECP-14,' 'SECP-15,' 'SECP-16,' 'SECP-17,' 'SECP-18,' 'SECP-19,' 'SECP-20,' 'SECP-21,' 'SECP-22,' 'SECP-23,' 'SECP-24,' 'SECP-25,' 'SECP-26,' 'SECP-27,' 'SECP-28,' 'SECP-29,' 'SECP-30,' 'SECP-31,' 'SECP-32,' 'SECP-33,' 'SECP-34,' 'SECP-35,' 'SECP-36,' 'SECP-37,' 'SECP-38,' 'SECP-39,' 'SECP-40,' 'SECP-41,' 'SECP-42,' 'SECP-43,' 'SECP-44,' 'SECP-45,' 'SECP-46,' 'SECP-47,' 'SECP-48,' 'SECP-49,' 'SECP-50,' 'SECP-51,' 'SECP-52,' 'SECP-53,' 'SECP-54,' 'SECP-55,' 'SECP-56,' 'SECP-57,' 'SECP-58,' 'SECP-59,' 'SECP-60,' and 'SECP-61' and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for utilizing the purified secreted proteins and/or their encoding polynucleotides for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified secreted proteins and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-61.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the

group consisting of SEQ ID NO:1-61. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-61. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:62-122.

Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61.

Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group

consisting of SEQ ID NO:62-122, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:62-122, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:62-122, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:62-122, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:62-122, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:62-122, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, b) a polypeptide comprising a

naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-61. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional SECP, comprising administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional SECP, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional

SECP, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-61. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:62-122, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said

method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:62-122, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:62-122, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:62-122, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:62-122, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"SECP" refers to the amino acid sequences of substantially purified SECP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of SECP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of SECP either by directly interacting with SECP or by acting on components of the biological pathway in which SECP participates.

An "allelic variant" is an alternative form of the gene encoding SECP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding SECP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as SECP or a polypeptide with at least one functional characteristic of SECP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding SECP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding SECP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent SECP. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of SECP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of SECP. Antagonists may include proteins such as antibodies, anticalins, nucleic acids,

carbohydrates, small molecules, or any other compound or composition which modulates the activity of SECP either by directly interacting with SECP or by acting on components of the biological pathway in which SECP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind SECP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or $2'\text{-}NH_2$), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed

nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic SECP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding SECP or fragments of SECP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (Accelrys,

Burlington MA) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a

measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of SECP or a polynucleotide encoding SECP which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:62-122 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:62-122, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:62-122 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:62-122 from related polynucleotides. The precise length of a fragment of SEQ ID NO:62-122 and the region of SEQ ID NO:62-122 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-61 is encoded by a fragment of SEQ ID NO:62-122. A fragment of SEQ ID NO:1-61 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-61. For example, a fragment of SEQ ID NO:1-61 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-61. The precise length of a fragment of SEQ ID NO:1-61 and the region of SEQ ID NO:1-61 to which the

fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A "full length" polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, alternatively, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of identical nucleotide matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of identical residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. The phrases "percent similarity" and "% similarity," as applied to polypeptide sequences, refer to the percentage of residue matches, including identical residue matches and conservative substitutions, between at least two polypeptide sequences aligned using a standardized algorithm. In contrast, conservative substitutions are not included in the calculation of percent identity between polypeptide sequences.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap

penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas

wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY, ch. 9).

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect

cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of SECP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of SECP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of SECP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of SECP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an SECP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of SECP.

"Probe" refers to nucleic acids encoding SECP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical

labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in, for example, Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments,

thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook and Russell (*supra*). The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing SECP, nucleic acids encoding SECP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or

cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with

a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook and Russell (*supra*).

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity or sequence similarity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity or sequence similarity over a certain defined length of one

of the polypeptides.

THE INVENTION

Various embodiments of the invention include new human secreted proteins (SECP), the polynucleotides encoding SECP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, developmental, and metabolic disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

Table 2 shows sequences with homology to polypeptide embodiments of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Accelrys, Burlington MA).

Column 6 shows amino acid residues comprising signature sequences, domains, and motifs, including the locations of signal peptides (as indicated by "Signal Peptide" and/or "signal_cleavage"). Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are secreted proteins. For example, SEQ ID NO:1 is 82% identical, from residue M1 to residue G229, to mouse punc protein (GenBank ID g3068592) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 9.8e-101, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also has homology to proteins that are members of the immunoglobulin cell surface protein superfamily, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:1 also contains immunoglobulin domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM, SMART and INCY databases of conserved protein families/domains. (See Table 3.) Data from SPSCAN and HMMER analyses provide further corroborative evidence that SEQ ID NO:1 is a secreted protein.

In another example, SEQ ID NO:41 is 94% identical, from residue G34 to residue Q230, to *H. sapiens* Con1 (GenBank ID g1911490) as determined by BLAST. (See Table 2.) The BLAST probability score is 4.3e-111. SEQ ID NO:41 also has homology to proteins that are localized to the extracellular region, as well as homology to basic proline-rich salivary proteins secreted by the parotid glands, which are similar to members of the BstNI-type subfamily of the salivary proline-rich protein (PRP) family, as determined by BLAST analysis using the PROTEOME database. Data from BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:41 has a basic proline-rich domain.

In yet another example, SEQ ID NO:60 is 82% identical from residue M1 to K160, and 100% identical from residue T124 to residue S350 to human ADIR1 (GenBank ID g10862826) as determined by BLAST. (See Table 2.) The BLAST probability score is 5.5e-195. SEQ ID NO:60 also has homology to torsin proteins, which have functional domains similar to those found in the Hsp100/clp family of ATPase chaperones, and are expressed in the central nervous system, as determined by BLAST analysis using the PROTEOME database. Data from BLAST analyses against the PRODOM database, provides further corroborative evidence that SEQ ID NO:60 is a torsin. SEQ ID NO:2-40, SEQ ID NO:42-59, and SEQ ID NO:61 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-61 are described in Table 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:62-122 or that distinguish between SEQ ID NO:62-122 and related polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier

(denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected)

indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses SECP variants. Various embodiments of SECP variants can have at least about 80%, at least about 90%, or at least about 95% amino acid sequence identity to the SECP amino acid sequence, and can contain at least one functional or structural characteristic of SECP.

Various embodiments also encompass polynucleotides which encode SECP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:62-122, which encodes SECP. The polynucleotide sequences of SEQ ID NO:62-122, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding SECP. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding SECP. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:62-122 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:62-122. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of SECP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding SECP. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding SECP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding SECP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding SECP. For example, a polynucleotide comprising a sequence of SEQ ID NO:68 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:69, a polynucleotide comprising a sequence of SEQ ID NO:76 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:77, a polynucleotide comprising a sequence of SEQ ID NO:78 is a splice variant of a polynucleotide

comprising a sequence of SEQ ID NO:86, a polynucleotide comprising a sequence of SEQ ID NO:101 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:102, and a polynucleotide comprising a sequence of SEQ ID NO:104 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:112. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of SECP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding SECP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring SECP, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode SECP and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring SECP under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding SECP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding SECP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode SECP and SECP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding SECP or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:62-122 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding SECP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been

size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode SECP may be cloned in recombinant DNA molecules that direct expression of SECP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express SECP.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter SECP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of SECP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of

DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, polynucleotides encoding SECP may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, SECP itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of SECP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

In order to express a biologically active SECP, the polynucleotides encoding SECP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding SECP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding SECP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding SECP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation

codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding SECP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook and Russell, *supra*, ch. 1-4, and 8; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding SECP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook and Russell, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding SECP. For example, routine cloning, subcloning, and propagation of polynucleotides encoding SECP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding SECP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription,

dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of SECP are needed, e.g. for the production of antibodies, vectors which direct high level expression of SECP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of SECP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

Plant systems may also be used for expression of SECP. Transcription of polynucleotides encoding SECP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding SECP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses SECP in host cells (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression

of SECP in cell lines is preferred. For example, polynucleotides encoding SECP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ and *apr*$^-$ cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β-glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding SECP is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding SECP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding SECP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding SECP and that express SECP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based

technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of SECP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on SECP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding SECP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding SECP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding SECP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode SECP may be designed to contain signal sequences which direct secretion of SECP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or

activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding SECP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric SECP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of SECP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the SECP encoding sequence and the heterologous protein sequence, so that SECP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled SECP may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

SECP, fragments of SECP, or variants of SECP may be used to screen for compounds that specifically bind to SECP. One or more test compounds may be screened for specific binding to SECP. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to SECP. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of SECP can be used to screen for binding of test compounds, such as antibodies, to SECP, a variant of SECP, or a combination of SECP and/or one or

more variants SECP. In an embodiment, a variant of SECP can be used to screen for compounds that bind to a variant of SECP, but not to SECP having the exact sequence of a sequence of SEQ ID NO:1-61. SECP variants used to perform such screening can have a range of about 50% to about 99% sequence identity to SECP, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to SECP can be closely related to the natural ligand of SECP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor SECP (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to SECP can be closely related to the natural receptor to which SECP binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for SECP which is capable of propagating a signal, or a decoy receptor for SECP which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG_1 (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to SECP, fragments of SECP, or variants of SECP. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of SECP. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of SECP. In another embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of SECP.

In an embodiment, anticalins can be screened for specific binding to SECP, fragments of SECP, or variants of SECP. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel

having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit SECP involves producing appropriate cells which express SECP, either as a secreted protein or on the cell membrane. Preferred cells can include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing SECP or cell membrane fractions which contain SECP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either SECP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with SECP, either in solution or affixed to a solid support, and detecting the binding of SECP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

SECP, fragments of SECP, or variants of SECP may be used to screen for compounds that modulate the activity of SECP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for SECP activity, wherein SECP is combined with at least one test compound, and the activity of SECP in the

presence of a test compound is compared with the activity of SECP in the absence of the test compound. A change in the activity of SECP in the presence of the test compound is indicative of a compound that modulates the activity of SECP. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising SECP under conditions suitable for SECP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of SECP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding SECP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding SECP may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding SECP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding SECP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease.

Alternatively, a mammal inbred to overexpress SECP, e.g., by secreting SECP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of SECP and secreted proteins. In addition, examples of tissues expressing SECP can be found in Table 6 and can also be found in Example XI. Therefore, SECP appears to play a role in cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, developmental, and metabolic disorders. In the treatment of disorders associated with increased SECP expression or activity, it is desirable to decrease the expression or activity of SECP. In the treatment of disorders associated with decreased SECP expression or activity, it is desirable to increase the expression or activity of SECP.

Therefore, in one embodiment, SECP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of SECP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cardiovascular disorder such as

congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, complications of cardiac transplantation, arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss, and a metabolic disorder such as diabetes, obesity,

hypertension, and atherosclerosis.

In another embodiment, a vector capable of expressing SECP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of SECP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified SECP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of SECP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of SECP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of SECP including, but not limited to, those listed above.

In a further embodiment, an antagonist of SECP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of SECP. Examples of such disorders include, but are not limited to, those cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, developmental, and metabolic disorders described above. In one aspect, an antibody which specifically binds SECP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express SECP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding SECP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of SECP including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of SECP may be produced using methods which are generally known in the art. In particular, purified SECP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind SECP. Antibodies to SECP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. In an embodiment, neutralizing antibodies (i.e.,

those which inhibit dimer formation) can be used therapeutically. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have application in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with SECP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to SECP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are substantially identical to a portion of the amino acid sequence of the natural protein. Short stretches of SECP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to SECP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce SECP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte

population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for SECP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between SECP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering SECP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for SECP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of SECP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple SECP epitopes, represents the average affinity, or avidity, of the antibodies for SECP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular SECP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the SECP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of SECP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml,

preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of SECP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding SECP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding SECP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding SECP (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102:469-475; Scanlon, K.J. et al. (1995) 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) Br. Med. Bull. 51:217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87:1308-1315; Morris, M.C. et al. (1997) Nucleic Acids Res. 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding SECP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites

(e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in SECP expression or regulation causes disease, the expression of SECP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in SECP are treated by constructing mammalian expression vectors encoding SECP and introducing these vectors by mechanical means into SECP-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of SECP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). SECP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding SECP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of

these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to SECP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding SECP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding SECP to cells which have one or more genetic abnormalities with respect to the expression of SECP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; Annu. Rev. Nutr. 19:511-544) and Verma, I.M. and N. Somia (1997; Nature 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding SECP to target cells which have one or more genetic abnormalities with respect to the expression of SECP. The use of herpes simplex virus (HSV)-based vectors may be

especially valuable for introducing SECP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; J. Virol. 73:519-532) and Xu, H. et al. (1994; Dev. Biol. 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding SECP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for SECP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of SECP-coding RNAs and the synthesis of high levels of SECP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of SECP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding SECP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding SECP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine,

cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

In other embodiments of the invention, the expression of one or more selected polynucleotides of the present invention can be altered, inhibited, decreased, or silenced using RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) methods known in the art. RNAi is a post-transcriptional mode of gene silencing in which double-stranded RNA (dsRNA) introduced into a targeted cell specifically suppresses the expression of the homologous gene (i.e., the gene bearing the sequence complementary to the dsRNA). This effectively knocks out or substantially reduces the expression of the targeted gene. PTGS can also be accomplished by use of DNA or DNA fragments as well. RNAi methods are described by Fire, A. et al. (1998; Nature 391:806-811) and Gura, T. (2000; Nature 404:804-808). PTGS can also be initiated by introduction of a complementary segment of DNA into the selected tissue using gene delivery and/or viral vector delivery methods described herein or known in the art.

RNAi can be induced in mammalian cells by the use of small interfering RNA also known as siRNA. SiRNA are shorter segments of dsRNA (typically about 21 to 23 nucleotides in length) that result *in vivo* from cleavage of introduced dsRNA by the action of an endogenous ribonuclease. SiRNA appear to be the mediators of the RNAi effect in mammals. The most effective siRNAs appear to be 21 nucleotide dsRNAs with 2 nucleotide 3' overhangs. The use of siRNA for inducing RNAi in mammalian cells is described by Elbashir, S.M. et al. (2001; Nature 411:494-498).

SiRNA can either be generated indirectly by introduction of dsRNA into the targeted cell, or directly by mammalian transfection methods and agents described herein or known in the art (such as liposome-mediated transfection, viral vector methods, or other polynucleotide delivery/introductory methods). Suitable SiRNAs can be selected by examining a transcript of the target polynucleotide (e.g., mRNA) for nucleotide sequences downstream from the AUG start codon and recording the occurrence of each nucleotide and the 3' adjacent 19 to 23 nucleotides as potential siRNA target sites, with sequences having a 21 nucleotide length being preferred. Regions to be avoided for target siRNA sites include the 5' and 3' untranslated regions (UTRs) and regions near the start codon (within 75 bases), as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP endonuclease complex. The selected target sites for siRNA can then be compared to the appropriate genome database (e.g., human, etc.) using BLAST or other sequence comparison algorithms known in the art. Target sequences with significant homology to other coding sequences can be eliminated from consideration. The selected SiRNAs can be produced by chemical synthesis methods known in the art or by *in vitro* transcription using commercially available methods and kits such as the SILENCER siRNA

construction kit (Ambion, Austin TX).

In alternative embodiments, long-term gene silencing and/or RNAi effects can be induced in selected tissue using expression vectors that continuously express siRNA. This can be accomplished using expression vectors that are engineered to express hairpin RNAs (shRNAs) using methods known in the art (see, e.g., Brummelkamp, T.R. et al. (2002) Science 296:550-553; and Paddison, P.J. et al. (2002) Genes Dev. 16:948-958). In these and related embodiments, shRNAs can be delivered to target cells using expression vectors known in the art. An example of a suitable expression vector for delivery of siRNA is the PSILENCER1.0-U6 (circular) plasmid (Ambion). Once delivered to the target tissue, shRNAs are processed *in vivo* into siRNA-like molecules capable of carrying out gene-specific silencing.

In various embodiments, the expression levels of genes targeted by RNAi or PTGS methods can be determined by assays for mRNA and/or protein analysis. Expression levels of the mRNA of a targeted gene, can be determined by northern analysis methods using, for example, the NORTHERNMAX-GLY kit (Ambion); by microarray methods; by PCR methods; by real time PCR methods; and by other RNA/polynucleotide assays known in the art or described herein. Expression levels of the protein encoded by the targeted gene can be determined by Western analysis using standard techniques known in the art.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding SECP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased SECP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding SECP may be therapeutically useful, and in the treatment of disorders associated with decreased SECP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding SECP may be therapeutically useful.

In various embodiments, one or more test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound

based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding SECP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding SECP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding SECP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may

consist of SECP, antibodies to SECP, and mimetics, agonists, antagonists, or inhibitors of SECP.

In various embodiments, the compositions described herein, such as pharmaceutical compositions, may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery allows administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising SECP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, SECP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example SECP or fragments thereof, antibodies of SECP, and agonists, antagonists or inhibitors of SECP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the

therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind SECP may be used for the diagnosis of disorders characterized by expression of SECP, or in assays to monitor patients being treated with SECP or agonists, antagonists, or inhibitors of SECP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for SECP include methods which utilize the antibody and a label to detect SECP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring SECP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of SECP expression. Normal or standard values for SECP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to SECP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of SECP expressed in

subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding SECP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of SECP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of SECP, and to monitor regulation of SECP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding SECP or closely related molecules may be used to identify nucleic acid sequences which encode SECP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding SECP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the SECP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:62-122 or from genomic sequences including promoters, enhancers, and introns of the SECP gene.

Means for producing specific hybridization probes for polynucleotides encoding SECP include the cloning of polynucleotides encoding SECP or SECP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding SECP may be used for the diagnosis of disorders associated with expression of SECP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory

disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cardiovascular disorder such as congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, complications of cardiac transplantation, arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other

neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss, and a metabolic disorder such as diabetes, obesity, hypertension, and atherosclerosis. Polynucleotides encoding SECP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered SECP expression. Such qualitative or quantitative methods are well known in the art.

In a particular embodiment, polynucleotides encoding SECP may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding SECP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding SECP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of SECP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding SECP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from

samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding SECP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding SECP, or a fragment of a polynucleotide complementary to the polynucleotide encoding SECP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding SECP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding SECP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out

sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of SECP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and

effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, SECP, fragments of SECP, or antibodies specific for SECP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids

in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for SECP to quantify the levels of SECP expression. In one embodiment, the antibodies are used as elements on a microarray,

and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/25116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known

and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding SECP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding SECP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, SECP, its catalytic or immunogenic fragments, or

oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between SECP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with SECP, or fragments thereof, and washed. Bound SECP is then detected by methods well known in the art. Purified SECP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding SECP specifically compete with a test compound for binding SECP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with SECP.

In additional embodiments, the nucleotide sequences which encode SECP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/370,707, U.S. Ser. No. 60/373,824, U.S. Ser. No. 60/377,883, and U.S. Ser. No. 60/383,218, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated

from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen, Carlsbad CA), PCDNA2.1 plasmid (Invitrogen), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte, Palo Alto CA), pRARE (Incyte), or pINCY (Incyte), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically

using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA

assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:62-122. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative secreted proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for

Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode secreted proteins, the encoded polypeptides were analyzed by querying against PFAM models for secreted proteins. Potential secreted proteins were also identified by homology to Incyte cDNA sequences that had been annotated as secreted proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended

with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of SECP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:62-122 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:62-122 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

Association of SECP Polynucleotides with Parkinson's Disease

Several genes have been identified as showing linkage to autosomal dominant forms of Parkinson's Disease (PD). PD is a common neurodegenerative disorder causing bradykinesia, resting tremor, muscular rigidity, and postural instability. Cytoplasmic eosinophilic inclusions called Lewy bodies, and neuronal loss especially in the substantia nigra pars compacta, are pathological hallmarks of PD (Valente, E.M. et al (2001) Am. J. Hum. Genet. 68:895-900). Lewy body Parkinson disease has been thought to be a specific autosomal dominant disorder (Wakabayashi, K. et al. (1998) Acta Neuropath. 96:207-210). Juvenile parkinsonism may be a specific autosomal recessive disorder (Matsumine, H. et al. (1997) Am. J. Hum. Genet. 60: 588-596). (Online Mendelian Inheritance in Man, OMIM. Johns Hopkins University, Baltimore, MD. MIM Number: 168600: Sept. 9, 2002: . World Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/)

Association of a disease with a chromosomal locus can be determined by lod score. Lod score is a statistical method used to test the linkage of two or more loci within families having a genetic disease. The lod score is the logarithm to base 10 of the odds in favor of linkage. Linkage is defined as the tendency of two genes located on the same chromosome to be inherited together through meiosis (*Genetics in Medicine*, Fifth Edition, (1991) Thompson, M.W. et al., W.B. Saunders Co. Philadelphia). A lod score of +3 or greater (1000:1 odds in favor of linkage) indicates a probability of 1 in 1000 that a particular marker was found solely by chance in affected individuals, which is strong evidence that two genetic loci are linked.

One such gene implicated in PD is PARK3, which maps to 2p13 (Gasser, T. et al. (1998) Nature Genet. 18:262-265). A marker at chromosomal position D2S441 was found to have a lod score of 3.2 in the region of PARK3. This marker supported the disease association of PARK3 in the chromosomal interval from D2S134 to D2S286 (Gasser et al., *supra*). Markers located within chromosomal intervals D2S134 and D2S286, which map between 83.88 to 94.05 centiMorgans on the short arm of chromosome 2, were used to identify genes that map in the region between D2S134 and D2S286.

A second PD gene, implicated in early-onset recessive parkinsonism, is PARK6, located on chromosome 1 at 1p35-1p36. Several markers were obtained with lod scores greater than 3 including D1S199, D1S2732, D1S2828, D1S478, D1S2702, D1S2734, D1S2674 (Valente, E.M. et al, supra). These markers were used to determine the PD-relevant range of chromosome loci and identify sequences that map to chromosome 1 between D1S199 and D1S2885.

Restriction fragment length polymorphism (RFLP) markers shown to be near regions of DNA known as sequence-tagged sites (STS), have been mapped to NT_Contigs generated by the Human Genome Project using ePCR (Schuler, G.D. (1997) Genome Research 7: 541-550, and (1998) Trends

Biotechnol. 16(11):456-459). Contigs containing regions of DNA with known disease-associated markers are therefore used to identify SECP sequences that map to disease-associated regions of the genome.

Polynucleotides encoding SECP were mapped to NT_Contigs. Contigs longer than 1Mb were broken into subcontigs of 1Mb length with overlaping sections of 100kb. A preliminary step used an algorithm, similar to MEGABLAST, to define the mRNA sequence /masked genomic DNA contig pairings. The cDNA/genomic pairings identified by the first algorithm were confirmed, and the SECP polynucleotides mapped to DNA contigs, using SIM4 (Florea, L. et al. (1998) Genome Res. 8:967-974, version May 2000) which had been optimized for high throughput processing and strand assignment confidence. The SIM4 output of the mRNA sequence/genomic contig pairs was further processed to determine the correct location of the SECP polynucleotides on the genomic contig, as well as their strand identity.

SEQ ID NO:76 was mapped to Contig NT_005428 from Genbank, version 128, covering a 9.65 Mb region of the genome that also contains PD-associated genetic markers D2S134 and D2S286. The maximum distance between SEQ ID NO:76 and markers D2S134 and D2S286, therefore, is 9.65 Mb. Thus, SEQ ID NO:76 is in proximity with genetic markers shown to consistently associate with PD. Also, SEQ ID NO:77 was mapped to Contig NT_005428 from Genbank, version 128, covering a 9.65 Mb region of the genome that also contains PD-associated genetic markers D2S134 and D2S286. The maximum distance between SEQ ID NO:77 and markers D2S134 and D2S286, therefore, is 9.65 Mb. Thus, SEQ ID NO:77 is in proximity with genetic markers shown to consistently associate with PD. Therefore, in various embodiments, SEQ ID NO:76-77 can be used for one or more of the following: i) linkage analysis of persons and/or families to the PD disease region at 2p13, ii) diagnostic assays for PD, and iii) developing therapeutics and/or other treatments for PD.

SEQ ID NO:83 was mapped to Contig NT_004610 from Genbank, version 128, covering a 14.87 Mb region of the genome that also contains PD-associated genetic markers D1S199 and D1S2885. The maximum distance between SEQ ID NO:83 and markers D1S199 and D1S2885, therefore, is 14.87 Mb. Thus, SEQ ID NO:83 is in proximity with genetic markers shown to consistently associate with PD. Therefore, in various embodiments, SEQ ID NO:83 can be used for one or more of the following: i) linkage analysis of persons and/or families to the PD disease region at 1p35-1p36, ii) diagnostic assays for PD, and iii) developing therapeutics and/or other treatments for PD.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook and Russell, *supra*, ch. 7; Ausubel et

al., *supra*, ch. 4).

Analogous computer techniques applying BLAST were used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \times \text{minimum } \{\text{length(Seq. 1), length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding SECP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided

by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding SECP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte, Palo Alto CA).

VIII. Extension of SECP Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and

sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in SECP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:62-122 using the LIFESEQ database (Incyte). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in

immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:62-122 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed.

(1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (Clontech, Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma-Aldrich, St. Louis MO) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines

at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte). Array elements that exhibit at least about a two-fold change in expression, a signal-to-background ratio of at least about 2.5, and an element spot size of at least about 40%, are considered to be

differentially expressed.

Expression

For example, SEQ ID NO:66 showed differential expression associated with colon cancer, as determined by microarray analysis. Gene expression profiles of the following tissues from the same donor were compared: normal sigmoid colon tissue and sigmoid colon tumor tissue originating from a metastatic gastric sarcoma (stromal tumor) from a 48-year-old female (Huntsman Cancer Institute, Salt Lake City, UT); normal colon tissue and moderately differentiated colon adenocarcinoma from a 67-year-old donor; normal colon tissue and colon tumor tissue from an 83-year-old donor. The expression of SEQ ID NO:66 was downregulated by at least two-fold in the colon tumor tissue of each donor as compared to the normal colon tissue.

Further, SEQ ID NO:83 showed differential expression associated with colon cancer, as determined by microarray analysis. Gene expression profiles were obtained by comparing tumorous rectal tissue to normal colon tissue from the same donor, a 38-year-old male with invasive, poorly differentiated adenocarcinoma with metastases to 2 out of 13 lymph nodes surveyed (Huntsman Cancer Institute, Salt Lake City, UT; TNM classification: T3, N1, Mx). Different pieces of normal tissue were also compared against a pool of normal tissue from the same donor to determine gene expression variation in normal colon tissue. SEQ ID NO:83 showed at least two-fold decreased gene expression in diseased versus normal tissue from the same donor. Also, SEQ ID NO:83 showed at least two-fold decreased expression in diseased versus normal colon tissue when gene expression in normal colon tissue (mRNA pooled from 3 different donors) was compared to colon polyps from 2 different donors and colon tumors from 3 different donors.

Yet further, SEQ ID NO:101, SEQ ID NO:102, and SEQ ID NO:105 showed differential expression in association with colon cancer, as determined by microarray analysis. Gene expression profiles were obtained by comparing the results of competitive hybridization experiments. The gene expression profiles of 6 different colon cancer tissues were analyzed by comparing one individual sample to 5 others, keeping one element in common between the various pairs of comparisons. The reference tissue sample is a metastatic adenocarcinoma of ovarian origin, which distinguishes this sample from the others and may be of special interest. The other five samples include tumorous colon tissue collected from an 85-year-old male, an 81-year-old male, an 83-year-old female, as well as a mucinous adenocarcinoma from a 58-year-old female, and a poorly differentiated metastatic adenocarcinoma from a 56-year-old female. The gene expression of SEQ ID NO:101, SEQ ID NO:102, and SEQ ID NO:105 was decreased by two-fold in the tumorous rectal tissue samples as compared to the reference tissue. Therefore, SEQ ID NO:101, SEQ ID NO:102, and SEQ ID NO:105 may be useful as diagnostic markers for colon cancer, as well as for monitoring the progression and

treatment of colon cancer.

Also, SEQ ID NO:122 showed differential expression in association with colon cancer, as determined by microarray analysis. Gene expression profiles were obtained by comparing the results of competitive hybridization experiments between normal colon tissue and tumorous rectal tissue from the same donor. Tissue samples were collected from tumorous colon tissue from eight different donors (Huntsman Cancer Institute, Salt Lake City, UT). In eight separate matched tissue experiments, the expression of SEQ ID NO:122 was decreased between two- and eight-fold in the tumorous colon tissue as compared to grossly uninvolved colon tissue samples originating from the mathed donors. Thus, in an embodiment, SEQ ID NO:122 can be used in diagnostic assays for, and/or for monitoring the diagnosis, prognosis or treatment of colon cancer. Therefore, in various embodiments, SEQ ID NO:66, SEQ ID NO:83, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:105, and SEQ ID NO:122 can be used for one or more of the following: i) monitoring treatment of colon cancer, ii) diagnostic assays for colon cancer, and iii) developing therapeutics and/or other treatments for colon cancer.

In another example, SEQ ID NO:66, SEQ ID NO:70, and SEQ ID NO:91 showed differential expression associated with lung cancer, as determined by microarray analysis. The following pair comparisons were carried out: grossly uninvolved lung tissue from a 66 year-old male was compared to lung squamous cell carcinoma tissue from the same donor (Roy Castle International Centre for Lung Cancer Research, Liverpool, UK); normal lung was compared to lung tumor from the same donor, a 75 year-old female; and normal lung tissue from a 68 year-old female was compared to lung tumor from the same donor (Roy Castle International Centre for Lung Cancer Research, Liverpool, UK). The expression of SEQ ID NO:66 and SEQ ID NO:70 was upregulated by at least two-fold in all tumor tissues examined as compared to normal tissues.

Also, SEQ ID NO:91 showed differential expression in association with lung cancer. In three separate experiments, normal lung tissue was compared to lung tumor tissue from the same donor; in the first case, a 73 year-old male, in the second, a 66 year-old male, and in the third, a 68 year-old female. Gene expression of SEQ ID NO:91 was decreased by at least two-fold in lung tumor tissue versus normal lung tissue.

Further, SEQ ID NO:121 and SEQ ID NO:122 showed differential expression in association with lung cancer. Gene expression profiles were obtained by comparing the results of competitive hybridization experiments. Messenger RNA isolated from grossly uninvolved lung tissue with no visible abnormalities was compared to lung squamous cell adenocarcinoma tissue from the same donor (Roy Castle International Centre for Lung Cancer Research, Liverpool, UK). In two separate matched tissue experiments, the expression of SEQ ID NO:121 was increased by at least two-fold in

tumorous lung tissue as compared to normal lung tissue from the same donor. In three separate matched tissue experiments, the expression of SEQ ID NO:122 was decreased by at least two-fold in tumorous lung tissue as compared to normal lung tissue from the same donor. Therefore, in various embodiments, SEQ ID NO:66, SEQ ID NO:70, SEQ ID NO:91, SEQ ID NO:121, and SEQ ID NO:122 can be used for one or more of the following: i) monitoring treatment of lung cancer, ii) diagnostic assays for lung cancer, and iii) developing therapeutics and/or other treatments for lung cancer.

In another example, SEQ ID NO:78 and SEQ ID NO:86 showed differential expression associated with prostate cancer cell lines. The gene expression profiles of prostate carcinoma lines representative of the different stages of tumor progression were compared to that of prostate epithelial cells (PrECs). Cell lines compared included: a)PrEC, a primary prostate epithelial cell line isolated from a normal donor, b)DU 145, a prostate carcinoma cell line isolated from a metastatic site in the brain of 69-year old male with widespread metastatic prostate carcinoma, c)LNCaP, a prostate carcinoma cell line isolated from a lymph node biopsy of a 50-year-old male with metastatic prostate carcinoma (LNCaP cells are responsive to 5-alpha-dihydrotestosterone and express androgen receptors), and d)PC-3, a prostate adenocarcinoma cell line isolated from a metastatic site in the bone of a 62- year-old male with grade IV prostate adenocarcinoma. Cells grown under optimal conditions were compared to normal PrECs grown under restrictive conditions. The expression of SEQ ID NO:78 was upregulated by at least two-fold in DU145 and LNCaP cells as compared to PrEC cells. Also, differential expression of SEQ ID NO:86 was at least two-fold greater in diseased cells when compared to normal cells. Therefore, in various embodiments, SEQ ID NO:78 and SEQ ID NO:86 can be used for one or more of the following: i) monitoring treatment of prostate cancer, ii) diagnostic assays for prostate cancer, and iii) developing therapeutics and/or other treatments for prostate cancer.

In another example, SEQ ID NO:80 showed differential expression associated with ovarian cancer. A normal ovary from a 79 year-old female donor was compared to an ovarian tumor from the same donor (Huntsman Cancer Institute, Salt Lake City, UT). The expression of SEQ ID NO:80 was downregulated by at least two-fold in the tumor tissue as compared to the normal tissue. Therefore, in various embodiments, SEQ ID NO:80 can be used for one or more of the following: i) monitoring treatment of ovarian cancer, ii) diagnostic assays for ovarian cancer, and iii) developing therapeutics and/or other treatments for ovarian cancer.

In a further example, SEQ ID NO:70 showed differential expression associated with inflammatory immune responses, as determined by microarray analysis. PMA-activated THP-1 cells (monocytic) and untreated THP-1 cells (promonocytic) were stimulated *in vitro* with LPS for 4 hours.

LPS-treated THP-1 cells were compared to untreated THP-1 cells. The expression of SEQ ID NO:70 was downregulated by at least two-fold in LPS-stimulated monocytic (PMA-activated) THP-1 cells as compared to untreated promonocytic THP-1 cells. Therefore, in various embodiments, SEQ ID NO:70 can be used for one or more of the following: i) monitoring treatment of immune disorders and related diseases and conditions, ii) diagnostic assays for immune disorders and related diseases and conditions, and iii) developing therapeutics and/or other treatments for immune disorders and related diseases and conditions.

In another example, SEQ ID NO:95 showed differential expression associated with Tangier disease, as determined by microarray analysis of normal versus Tangier disease derived fibroblasts. Human fibroblasts were obtained from skin explants from both normal subjects and two patients with homozygous Tangier disease. Cell lines were immortalized by transfection with human papillomavirus 16 genes E6 and E7 and a neomycin resistance selectable marker. Both types of cells were cultured in the presence of cholesterol. Tangier disease derived cells are shown to be deficient in an assay of apoA-I mediated tritiated cholesterol efflux. Differential expression of SEQ ID NO:95 was at least two-fold greater in Tangier disease derived cells when compared to cells from normal subjects. Therefore, in various embodiments, SEQ ID NO:95 can be used for one or more of the following: i) monitoring treatment of Tangier disease, ii) diagnostic assays for Tangier disease, and iii) developing therapeutics and/or other treatments for Tangier disease.

In another example, SEQ ID NO:89 showed differential expression in a cultured cell line associated with liver cancer, as determined by microaray analysis. The C3A cell line is a clonal derivative of the HepG2 hepatoma cell line isolated from a 15-year-old male with a liver tumor. C3A cells express both the insulin receptor and the insulin-like growth factor II receptor. C3A cells have many characteristics of primary human hepatocytes in culture: i) expression of insulin receptor and insulin-like growth factor II receptor; ii) secretion of a high ratio of serum albumin compared with α-fetoprotein iii) conversion of ammonia to urea and glutamine; iv) metabolism of aromatic amino acids; and v) proliferation in glucose-free and insulin-free medium. The C3A cell line is now well established as an *in vitro* model of the mature human liver (Mickelson et al. (1995) Hepatology 22:866-875; Nagendra et al. (1997) Am J Physiol 272:G408-G416). Early Confluent C3A cells were treated with betamethasone at 1, 10, and 100 μM for 1, 3, and 6 hours. Gene expression of SEQ ID NO:89 was increased at least two-fold in human C3A cells derived from a hepatoblastoma. Therefore, in various embodiments, SEQ ID NO:89 can be used for one or more of the following: i) monitoring treatment of liver cancer, ii) diagnostic assays for liver cancer, and iii) developing therapeutics and/or other treatments for liver cancer.

In another example, SEQ ID NO:105, SEQ ID NO:106 and SEQ ID NO:108 showed

differential expression in association with breast cancer, as determined by microarray analysis. Gene expression profiles were obtained by comparing the results of competitive hybridization experiments. The gene expression profile of a breast carcinoma cell line treated with Epidermal growth factor (EGF) was compared to the gene expression profile of untreated cells from the same breast carcinoma line. BT-20 is a breast carcinoma cell line derived *in vitro* from the cells emigrating out of thin slices of the tumor mass isolated from a 74-year-old female. EGF is a potent mitogen for normal as well as malignant mammary epithelial cells. In addition, EGF found to induce tumor progression, and is highly expressed in breast carcinoma cells. EGF produces its effect by binding to its receptor, the protein tyrosine kinase EGF receptor (EGFR) also known as erbB2. Ligation of EGF to the EGFR activates the tyrosine kinase domain of EGFR and the phosphorylation of multiple substrates, thereby regulating gene expression. The expression of SEQ ID NO:105 was decreased by at least two-fold in the breast carcinoma line treated with EGF as compared to untreated cells from the same breast carcinoma line. In another example, a tumor from the right breast was compared to grossly uninvolved breast tissue from the same donor, a 43-year-old female diagnosed with invasive lobular carcinoma (Huntsman Cancer Institute, Salt Lake City, UT) *in situ*. The tumor was described as well differentiated and metastatic to 2 out of 13 lymph nodes. SEQ ID NO:106 showed increased gene expression by four-fold in the tumorous tissue sample as compared to a uninvolved tissue sample from the same donor. In another example, the gene expression profile of primary mammary epithelial cells (HMEC) was compared to that of various breast carcinoma lines at different stages of tumor progression. Cell lines compared included: a) BT-20 is a breast carcinoma cell line derived *in vitro* from the cells emigrating out thin slices of the tumor mass isolated from a 74-year-old female, b) BT-474, a breast ductal carcinoma cell line that was isolated from a solid, invasive ductal carcinoma of the breast obtained from a 60-year-old woman, c) BT-483, a breast ductal carcinoma cell line that was isolated from a papillary invasive ductal tumor obtained from a 23-year-old normal, menstruating, parous female with a family history of breast cancer, d) Hs578T, a breast ductal carcinoma cell line that was isolated from a 74-year-old female with breast carcinoma, e) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, f) MCF-10A, a breast mammary gland (luminal ductal characteristics) cell line that was isolated from a 36-year-old woman with fibrocystic breast disease, and g) MDA-MB-468, a breast adenocarcinoma cell line isolated from the pleural effusion of a 51-year-old female with metastatic adenocarcinoma of the breast. The expression of SEQ ID NO:108 was decreased by at least two-fold in carcinoma cell lines BT20, BT474, BT483, Hs578T, MCF7, and MDA-MB-468 as compared to the HMEC population. Therefore, in various embodiments, SEQ ID NO:105, SEQ ID NO:106, and SEQ ID NO:108 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays

for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

In yet another example, SEQ ID NO:101 and SEQ ID NO:102 showed differential expression in association with bone cancer, as determined by microarray analysis. Gene expression profiles were obtained by comparing the results of competitive hybridization experiments. Messenger RNA from normal human osteoblasts was compared to mRNA from biopsy specimens, osteosarcoma tissues, or primary cultures or metastasized tissues. A normal osteoblast primary culture, NHOst 5488, was chosen to be the reference sample. One basic set of experiments is defined as the comparison of mRNA from biopsy specimen with that of definitive surgical specimen from the same patient. Extended study of this basic set includes mRNA from primary cell cultures of the definitive surgical specimen, muscle, or cartilage tissue from the same patient. Biopsy specimens, definitive surgical specimens, or lung metastatic tissues from different individuals were also included to reveal individual variability. The gene expression of SEQ ID NO:101 and SEQ ID NO:102 was decreased by at least two-fold in mRNA from biopsy specimens and osteosarcoma tissues as compared to the mRNA from the reference sample. Therefore, in various embodiments, SEQ ID NO:101 and SEQ ID NO:102 can be used for one or more of the following: i) monitoring treatment of bone cancer, ii) diagnostic assays for bone cancer, and iii) developing therapeutics and/or other treatments for bone cancer.

In still another example, SEQ ID NO:105 showed differential expression in association with inflammatory and immune responses, as determined by microarray analysis. Gene expression profiles were obtained by comparing the results of competitive hybridization experiments. Human peripheral blood mononuclelar cells (PBMCs) from seven healthy donors were stimulated *in vitro* with Staphylococal extoxin B (SEB) for 24 and 72 hours. The SEB treated PBMCs from each donor were compared to PBMCs from the same donor, kept in culture for 24 hours, in the absence of SEB. The gene expression of SEQ ID NO:16 was decreased by at least two-fold in SEB treated PBMCs as compared to untreated PBMCs from the same donors. In another example, PBMCs were collected from the blood of 6 healthy volunteer donors using standard gradient separation. The PBMCs from each donor were placed in culture for 2 and 4 hours in the presence of one or more cytokines selected from one of the two following cytokine groups: a) Cytokines associated positively with the inflammatory response (pro-inflammatory) such as IL-1 β , IL-2, IL-6, IL-8, IL-12, IL-18, IFN-γ , and TNF-α; or b) Cytokines negatively or neutrally associated with the inflammatory response (anti-inflammatory) such as IL-3, IL-4, IL-5, IL-7, IL-10, G-CSF, GM-CSF, Leptin, LIF, and TGF-β. Cytokine-treated PBMCs and untreated control-PBMCs from the different donors were pooled according to their respective treatments. The gene expression of SEQ ID NO:105 was decreased by at two-fold in PBMCs treated with both pro-inflammatory as well as anti-inflammatory cytokines, as

compared to untreated PBMCs. Therefore, in various embodiments, SEQ ID NO:105 can be used for one or more of the following: i) monitoring treatment of immune disorders and related diseases and conditions, ii) diagnostic assays for immune disorders and related diseases and conditions, and iii) developing therapeutics and/or other treatments for immune disorders and related diseases and conditions.

In a further example, SEQ ID NO:114 and SEQ ID NO:122 showed differential expression in association with metabolic disorders, as determined by microarray analysis. The expression of SEQ ID NO:114 was decreased by at least two-fold in human preadipocytes from an obese donor treated with a differentiation inducing medium when compared to non-treated preadipocytes from the same donor. The expression of SEQ ID NO:122 was decreased in preadipocytes from normal, overweight, and obese donors. The normal human primary subcutaneous preadipocytes were isolated from adipose tissue of a 28-year-old healthy female with a body mass index (BMI) of 23.59. The overweight human primary preadipocytes were isolated from adipose tissue of a 36-year-old overweight, but otherwise healthy female with a BMI of 27.7. The obese human primary preadipocytes were isolated from adipose tissue of a 40-year-old obese female with a BMI of 32.47. The preadipocytes were cultured and induced to differentiate into adipocytes by culturing them in the differentiation medium containing the active components, PPAR-γ agonist and human insulin. Human preadipocytes were treated with human insulin and PPAR-γ agonist for three days and subsequently were switched to medium containing insulin for 24 hours, 48 hours, 4 days, 8 days or 15 days before the cells were collected for analysis. Differentiated adipocytes were compared to untreated preadipocytes maintained in culture in the absence of inducing agents. Between 80% and 90% of the preadipocytes finally differentiated to adipocytes as observed under a phase contrast microscope. The experiments showed that the expression of SEQ ID NO:114 was decreased by at least two-fold in treated versus untreated preadipocytes from an obese donor, and that the expression of SEQ ID NO:122 was decreased by at least two-fold in treated versus untreated preadipocytes from normal, overweight, and obese donors. Therefore, in various embodiments, SEQ ID NO:114 and SEQ ID NO:122 can be used for one or more of the following: i) monitoring treatment of diabetes, obesity, hypertension, and atherosclerosis, ii) diagnostic assays for diabetes, obesity, hypertension, and atherosclerosis, and iii) developing therapeutics and/or other treatments for diabetes, obesity, hypertension, and atherosclerosis.

In addition, SEQ ID NO:118 showed tissue-specific expression. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta

(46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. The expression of SEQ ID NO:118 was increased by at least two-fold in the small intestine as compared to the reference sample. Therefore, in an embodiment, SEQ ID NO:118 can be used as a tissue marker for expression in the small intestine.

XII. Complementary Polynucleotides

Sequences complementary to the SECP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring SECP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of SECP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the SECP-encoding transcript.

XIII. Expression of SECP

Expression and purification of SECP is achieved using bacterial or virus-based expression systems. For expression of SECP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express SECP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of SECP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding SECP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, SECP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from SECP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified SECP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, XIX, and XX, where applicable.

XIV. Functional Assays

SECP function is assessed by expressing the sequences encoding SECP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

The influence of SECP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding SECP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding SECP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of SECP Specific Antibodies

SECP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the SECP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-SECP activity by, for example, binding the peptide or SECP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring SECP Using Specific Antibodies

Naturally occurring or recombinant SECP is substantially purified by immunoaffinity chromatography using antibodies specific for SECP. An immunoaffinity column is constructed by covalently coupling anti-SECP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing SECP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of SECP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/SECP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as

urea or thiocyanate ion), and SECP is collected.

XVII. Identification of Molecules Which Interact with SECP

SECP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled SECP, washed, and any wells with labeled SECP complex are assayed. Data obtained using different concentrations of SECP are used to calculate values for the number, affinity, and association of SECP with the candidate molecules.

Alternatively, molecules interacting with SECP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

SECP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of SECP Activity

An assay for growth stimulating or inhibiting activity of SECP measures the amount of DNA synthesis in Swiss mouse 3T3 cells (McKay, I. and I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York, NY). In this assay, varying amounts of SECP are added to quiescent 3T3 cultured cells in the presence of [^{3}H]thymidine, a radioactive DNA precursor. SECP for this assay can be obtained by recombinant means or from biochemical preparations. Incorporation of [^{3}H]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold SECP concentration range is indicative of growth modulating activity. One unit of activity per milliliter is defined as the concentration of SECP producing a 50% response level, where 100% represents maximal incorporation of [^{3}H]thymidine into acid-precipitable DNA .

Alternatively, an assay for SECP activity measures the stimulation or inhibition of neurotransmission in cultured cells. Cultured CHO fibroblasts are exposed to SECP. Following endocytic uptake of SECP, the cells are washed with fresh culture medium, and a whole cell voltage-clamped *Xenopus* myocyte is manipulated into contact with one of the fibroblasts in SECP-free medium. Membrane currents are recorded from the myocyte. Increased or decreased current relative to control values are indicative of neuromodulatory effects of SECP (Morimoto, T. et al. (1995) Neuron 15:689-696).

Alternatively, an assay for SECP activity measures the amount of SECP in secretory, membrane-bound organelles. Transfected cells as described above are harvested and lysed. The lysate is fractionated using methods known to those of skill in the art, for example, sucrose gradient ultracentrifugation. Such methods allow the isolation of subcellular components such as the Golgi apparatus, ER, small membrane-bound vesicles, and other secretory organelles. Immunoprecipitations from fractionated and total cell lysates are performed using SECP-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The concentration of SECP in secretory organelles relative to SECP in total cell lysate is proportional to the amount of SECP in transit through the secretory pathway.

Alternatively, AMP binding activity is measured by combining SECP with ^{32}P-labeled AMP. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to remove unbound label. The radioactivity retained in the gel is proportional to SECP activity.

XIX. Demonstration of Immunoglobulin Activity

An assay for SECP activity measures the ability of SECP to recognize and precipitate antigens from serum. This activity can be measured by the quantitative precipitin reaction. (Golub, E.S. et al. (1987) Immunology: A Synthesis, Sinauer Associates, Sunderland, MA, pp. 113-115.) SECP is isotopically labeled using methods known in the art. Various serum concentrations are added to constant amounts of labeled SECP. SECP-antigen complexes precipitate out of solution and are collected by centrifugation. The amount of precipitable SECP-antigen complex is proportional to the amount of radioisotope detected in the precipitate. The amount of precipitable SECP-antigen complex is plotted against the serum concentration. For various serum concentrations, a characteristic precipitin curve is obtained, in which the amount of precipitable SECP-antigen complex initially increases proportionately with increasing serum concentration, peaks at the equivalence point, and then decreases proportionately with further increases in serum concentration. Thus, the amount of precipitable SECP-antigen complex is a measure of SECP activity which is characterized by sensitivity to both limiting and excess quantities of antigen.

Alternatively, an assay for SECP activity measures the expression of SECP on the cell surface. cDNA encoding SECP is transfected into a non-leukocytic cell line. Cell surface proteins are labeled with biotin (de la Fuente, M.A. et al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using SECP-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of SECP expressed on the cell surface.

Alternatively, an assay for SECP activity measures the amount of cell aggregation induced by

overexpression of SECP. In this assay, cultured cells such as NIH3T3 are transfected with cDNA encoding SECP contained within a suitable mammalian expression vector under control of a strong promoter. Cotransfection with cDNA encoding a fluorescent marker protein, such as Green Fluorescent Protein (CLONTECH), is useful for identifying stable transfectants. The amount of cell agglutination, or clumping, associated with transfected cells is compared with that associated with untransfected cells. The amount of cell agglutination is a direct measure of SECP activity.

XX. SECP Secretion Assay

A high throughput assay may be used to identify polypeptides that are secreted in eukaryotic cells. In an example of such an assay, polypeptide expression libraries are constructed by fusing 5'-biased cDNAs to the 5'-end of a leaderless β-lactamase gene. β-lactamase is a convenient genetic reporter as it provides a high signal-to-noise ratio against low endogenous background activity and retains activity upon fusion to other proteins. A dual promoter system allows the expression of β-lactamase fusion polypeptides in bacteria or eukaryotic cells, using the *lac* or CMV promoter, respectively.

Libraries are first transformed into bacteria, *e.g.*, *E. coli*, to identify library members that encode fusion polypeptides capable of being secreted in a prokaryotic system. Mammalian signal sequences direct the translocation of β-lactamase fusion polypeptides into the periplasm of bacteria where they confer antibiotic resistance to carbenicillin. Carbenicillin-selected bacteria are isolated on solid media, individual clones are grown in liquid media, and the resulting cultures are used to isolate library member plasmid DNA.

Mammalian cells, *e.g.*, 293 cells, are seeded into 96-well tissue culture plates at a density of about 40,000 cells/well in 100 μl phenol red-free DME supplemented with 10% fetal bovine serum (FBS) (Life Technologies, Rockville, MD). The following day, purified plasmid DNAs isolated from carbenicillin-resistant bacteria are diluted with 15 μl OPTI-MEM I medium (Life Technologies) to a volume of 25 μl for each well of cells to be transfected. In separate plates, 1 μl LF2000 Reagent (Life Technologies) is diluted into 25 μl/well OPTI-MEM I. The 25 μl diluted LF2000 Reagent is then combined with the 25 μl diluted DNA, mixed briefly, and incubated for 20 minutes at room temperature. The resulting DNA-LF2000 reagent complexes are then added directly to each well of 293 cells. Cells are also transfected with appropriate control plasmids expressing either wild-type β-lactamase, leaderless β-lactamase, or, for example, CD4-fused leaderless β-lactamase. 24 hrs following transfection, about 90 μl of cell culture media are assayed at 37°C with 100 μM Nitrocefin (Calbiochem, San Diego, CA) and 0.5 mM oleic acid (Sigma Corp. St. Louis, MO) in 10 mM phosphate buffer (pH 7.0). Nitrocefin is a substrate for β-lactamase that undergoes a noticeable color change from yellow to red upon hydrolysis. β-lactamase activity is monitored over 20 min in a

microtiter plate reader at 486 nm. Increased color absorption at 486 nm corresponds to secretion of a β-lactamase fusion polypeptide in the transfected cell media, resulting from the presence of a eukaryotic signal sequence in the fusion polypeptide. Polynucleotide sequence analysis of the corresponding library member plasmid DNA is then used to identify the signal sequence-encoding cDNA. (Described in U.S. Patent application 09/803,317, filed March 9, 2001.)

For example, SEQ ID NO:6 was shown to be a secreted protein using this assay.

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7510186	1	7510186CD1	62	7510186CB1	
7510045	2	7510045CD1	63	7510045CB1	
7510258	3	7510258CD1	64	7510258CB1	56010433CA2, 56010441CA2, 56010449CA2, 56010541CA2, 56010549CA2, 56010557CA2, 90135770CA2, 90135786CA2
7510450	4	7510450CD1	65	7510450CB1	4896736CA2
7504544	5	7504544CD1	66	7504544CB1	2294047CA2, 3818522CA2, 5264426CA2
7504804	6	7504804CD1	67	7504804CB1	
7510156	7	7510156CD1	68	7510156CB1	90067137CA2, 90067221CA2, 95178372CA2
7510157	8	7510157CD1	69	7510157CB1	
7510993	9	7510993CD1	70	7510993CB1	4852722CA2
7511149	10	7511149CD1	71	7511149CB1	2859142CA2
7511184	11	7511184CD1	72	7511184CB1	3282234CA2
7511240	12	7511240CD1	73	7511240CB1	7408230CA2, 95228889CA2
7511376	13	7511376CD1	74	7511376CB1	90138321CA2, 90138493CA2
7501330	14	7501330CD1	75	7501330CB1	
7509961	15	7509961CD1	76	7509961CB1	2061402CA2, 7272262CA2, 95116310CA2, 95116374CA2
7509963	16	7509963CD1	77	7509963CB1	
7505089	17	7505089CD1	78	7505089CB1	
7510139	18	7510139CD1	79	7510139CB1	
7505053	19	7505053CD1	80	7505053CB1	
7511116	20	7511116CD1	81	7511116CB1	2516129CA2
7511175	21	7511175CD1	82	7511175CB1	3182329CA2
7504660	22	7504660CD1	83	7504660CB1	3511453CA2, 95124552CA2, 95124712CA2
7504681	23	7504681CD1	84	7504681CB1	90022239CA2
7506472	24	7506472CD1	85	7506472CB1	
7506483	25	7506483CD1	86	7506483CB1	2693149CA2
7506525	26	7506525CD1	87	7506525CB1	
7506549	27	7506549CD1	88	7506549CB1	
7506683	28	7506683CD1	89	7506683CB1	

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7510814	29	7510814CD1	90	7510814CB1	90023401CA2
7504727	30	7504727CD1	91	7504727CB1	95083144CA2, 95120469CA2, 95120545CA2
7506958	31	7506958CD1	92	7506958CB1	
7505332	32	7505332CD1	93	7505332CB1	
7505364	33	7505364CD1	94	7505364CB1	6827470CA2
7505455	34	7505455CD1	95	7505455CB1	
7505641	35	7505641CD1	96	7505641CB1	90032040CA2, 90032048CA2
7511242	36	7511242CD1	97	7511242CB1	6981966CA2, 90069104CA2, 90069112CA2, 90069120CA2, 90069128CA2, 90069144CA2, 90069204CA2, 90069212CA2, 90069220CA2, 90069228CA2, 90069236CA2, 90069244CA2, 90069339CA2, 90069486CA2, 90071613CA2, 90071621CA2, 90071629CA2, 90071645CA2, 90071705CA2, 90071713CA2, 90071721CA2, 90071729CA2
7506564	37	7506564CD1	98	7506564CB1	95137973CA2, 95138010CA2
7509076	38	7509076CD1	99	7509076CB1	
7506666	39	7506666CD1	100	7506666CB1	90119704CA2
7511731	40	7511731CD1	101	7511731CB1	
7511735	41	7511735CD1	102	7511735CB1	
7511729	42	7511729CD1	103	7511729CB1	1321834CA2
7511255	43	7511255CD1	104	7511255CB1	
7511628	44	7511628CD1	105	7511628CB1	
7511657	45	7511657CD1	106	7511657CB1	8551147CA2
7512343	46	7512343CD1	107	7512343CB1	90079786CA2, 90205575CA2
7512357	47	7512357CD1	108	7512357CB1	
7511046	48	7511046CD1	109	7511046CB1	1250593CA2, 1291101CA2, 1438315CA2, 90106103CA2, 90106111CA2, 90106119CA2, 90106127CA2, 90106211CA2, 90106219CA2, 90106227CA2, 90106235CA2, 90106243CA2
7512332	49	7512332CD1	110	7512332CB1	90085104CA2, 90085212CA2
7511219	50	7511219CD1	111	7511219CB1	3840272CA2, 90210964CA2, 90211064CA2

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7510933	51	7510933CD1	112	7510933CB1	7190894CA2
7511461	52	7511461CD1	113	7511461CB1	
7511808	53	7511808CD1	114	7511808CB1	5687043CA2
7511817	54	7511817CD1	115	7511817CB1	5422110CA2
7511832	55	7511832CD1	116	7511832CB1	2599523CA2
7512301	56	7512301CD1	117	7512301CB1	4113968CA2
7512320	57	7512320CD1	118	7512320CB1	90024756CA2, 90024772CA2, 90044608CA2, 90044624CA2, 90044632CA2
7512371	58	7512371CD1	119	7512371CB1	2207655CA2
7512442	59	7512442CD1	120	7512442CB1	90064962CA2
7512311	60	7512311CD1	121	7512311CB1	
7512474	61	7512474CD1	122	7512474CB1	

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	7510186CD1	g3068592	9.8E-101	[Mus musculus] punc Salbaum, J. M. Punc, a novel mouse gene of the immunoglobulin superfamily, is expressed predominantly in the developing nervous system Mech. Dev. 71, 201-204 (1998)
		582679|Punc	8.4E-102	[Mus musculus][Adhesin/agglutinin][Plasma membrane] Putative neuronal cell adhesion molecule, a member of the immunoglobulin superfamily of cell surface proteins, may play a role in cerebellar control of motor coordination and early embryogenesis Yang, W. et al. Impaired motor coordination in mice that lack punc. Mol Cell Biol 21, 6031-6043. (2001).
		735342|ROBO2	3.9E-24	[Homo sapiens][Receptor (signalling)] Roundabout (axon guidance receptor, Drosophila) homolog 2, a member of the immunoglobulin superfamily that is involved in axon guidance Kidd, T. et al. Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors. Cell 92, 205-215 (1998).
6	7504804CD1	598754|FLJ10890	0.0	[Homo sapiens] Protein containing four TPR (tetratrico peptide repeat) domains, which may mediate protein-protein interactions
14	7501330CD1	g1654000	2.8E-23	[Homo sapiens] Not56-like protein
		342664|NOT56L	2.2E-24	[Homo sapiens] Protein with low similarity to S. cerevisiae RHK1, which is a mannosyltransferase that is involved in N-glycosylation and in the synthesis of the HM-1 killer toxin receptor
15	7509961CD1	g5410294	1.1E-73	[Homo sapiens] ancient ubiquitous protein AUP1 isoform
		432680|AUP1	8.6E-75	[Homo sapiens] Ubiquitously expressed protein Jang, W. et al. Aup1, a novel gene on mouse chromosome 6 and human chromosome 2p13. Genomics 36, 366-368 (1996).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		580927|Aup1	4.1E-52	[Mus musculus] Ancient ubiquitous protein 1, has lipid binding motif and is expressed in all tissues; in humans, the human AUP1 gene is localized in the Parkinson's disease susceptibility locus Jang, W. et al. (supra)
16	7509963CD1	g5410292	1.9E-131	[Homo sapiens] ancient ubiquitous 46 kDa protein AUP1
		432680|AUP1	1.2E-130	[Homo sapiens] Ubiquitously expressed protein Jang, W. et al. (supra)
		580927|Aup1	9.8E-122	[Mus musculus] Ancient ubiquitous protein 1, has lipid binding motif and is expressed in all tissues; in humans, the human AUP1 gene is localized in the Parkinson's disease susceptibility locus Jang, W. et al. (supra)
17	7505089CD1	g6688145	2.6E-100	[Homo sapiens] NICE-3 protein Marenholz I, I. et al. Identification of Human Epidermal Differentiation Complex (EDC)-Encoded Genes by Subtractive Hybridization of Entire YACs to a Gridded Keratinocyte cDNA Library Genome Res. 11, 341-355 (2001)
		g14249862	3.10E-85	[Homo sapiens] DKFZP586G1722 protein
18	7510139CD1	g14714617	7.2E-139	[Homo sapiens] brain-specific membrane-anchored protein
		428694|BSMAP	6.0E-140	[Homo sapiens] Brain-specific membrane-anchored protein, a putative type-I membrane glycoprotein that may have a role in CNS function Elson, G. C. et al. BSMAP, a novel protein expressed specifically in the brain whose gene is localized on chromosome 19p12. Biochem Biophys Res Commun 264, 55-62 (1999).
19	7505053CD1	g3668141	5.5E-11	[Homo sapiens] PBK1 protein Huch, G. et al. Identification of differentially expressed genes in human trophoblast cells by differential-display RT-PCR. Placenta 19, 557-567 (1998)
20	7511116CD1	g1216486	4.8E-53	[Cricetulus griseus] HT protein

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		242468|F09E8.2	1.0E-41	[Caenorhabditis elegans] Protein containing EGF-like repeats, has weak similarity to human low density lipoprotein receptors and D. melanogaster TEN-1 (tenascin)
		g23272252	1.00E-121	cysteine-rich with EGF-like domains 1 [Mus musculus]
21	7511175CD1	g12803257	8.4E-40	[Homo sapiens] putative T1/ST2 receptor binding protein
		475543|LOC5099 9	5.3E-84	[Homo sapiens] Member of the emp24/gp25L/p24 family, which are implicated in transport of proteins from the ER to the Golgi
		428186|IL1RL1L G	7.2E-41	[Homo sapiens][Ligand] T1-ST2 receptor binding protein, binds to the T1/ST2 receptor (IL1RL1), predicted to be involved in signal transduction Gayle, M. A. et al. Cloning of a putative ligand for the T1/ST2 receptor. J Biol Chem 271, 5784-5789 (1996).
		319200|Il1rl1	1.9E-40	[Mus musculus][Ligand] Interleukin 1 receptor-like 1 ligand, binds to the T1/ST2 receptor (Il1rl1), predicted to be involved in signal transduction Gayle, M. A. et al. (supra)
23	7504681CD1	g24559836	6.00E-22	polycystic kidney disease 1-related protein [Homo sapiens]
26	7506525CD1	g22134528	3.00E-25	putative anion transporter [Homo sapiens]
27	7506549CD1	731697 |FLJ14735	1.9E-12	[Homo sapiens] Protein containing a low-density lipoprotein (LDL) receptor class A domain, has low similarity to serine protease inhibitor (Kunitz type 1, human SPINT1), which inhibits hepatocyte growth factor activator.
37	7506564CD1	g5916203	9.7E-78	[Rattus norvegicus] estrogen-regulated protein
		757628 |Itmap1	2.4E-79	[Rattus norvegicus][Receptor (protein translocation)][Plasma membrane] Estrogen regulated protein 1, a putative transmembrane protein that contains a zona pellucida binding domain and a CUB domain, may function in the reproductive cycle and early pregnancy, is tightly regulated by estrogen in the uterus and oviduct. Chen, D. et al. Cloning and uterus/oviduct-specific expression of a novel estrogen-regulated gene (ERG1) [published erratum appears in J Biol Chem 2000 Feb 18;275(7):5248]. J Biol Chem 274, 32215-24 (1999).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		437256 |PCOLCE2	5.9E-17	[Homo sapiens] Procollagen C-endopeptidase enhancer 2, a protein with moderate similarity to PCOLCE, a glycoprotein that enhances cleavage of type I procollagen C-propeptide; gene is a candidate disease gene for open-angle glaucoma. Xu, H. et al. Identification and Expression of a Novel Type I Procollagen C-Proteinase Enhancer Protein Gene from the Glaucoma Candidate Region on 3q21-q24. Genomics 66, 264-273 (2000).
38	7509076CD1	g1841553	4.5E-128	[Homo sapiens] NG3
	7509076CD1	697430 |NG3	3.8E-129	[Homo sapiens] Protein containing two epidermal growth factor (EGF)-like domains, has a region of moderate similarity to a region of fibrillin 2 (human FBN2), which is an extracellular matrix protein that regulates formation of connective tissue microfibrils.
		368734 |Mm.45990	1.1E-101	[Mus musculus] Protein containing EGF-like domains, which are found in some secreted proteins and extracellular domains of transmembrane proteins.
39	7506666CD1	g338075	3E-154	[Homo sapiens] sex hormone-binding globulin Hammond, G. L. et al. The human sex hormone-binding globulin gene contains exons for androgen-binding protein and two other testicular messenger RNAs. Mol. Endocrinol. 3, 1869-1876 (1989).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		568830 |SHBG	2.6E-155	[Homo sapiens][Structural protein; Small molecule-binding protein] [Extracellular (excluding cell wall)] Sex hormone binding globulin (androgen binding protein), a glycoprotein that binds and transports androgens and leads to protein kinase A (PRKACB) activation; plasma levels are associated with obesity, heart disease, and noninsulin-dependent diabetes. Misao, R., Nakanishi, Y., Fujimoto, J., Iwagaki, S., and Tamaya, T. Dominant expression of sex-hormone-binding-globulin exon-7 splicing variant over wild-type mRNA in human ovarian cancers. Int J Cancer 77, 828-32 (1998). Gascon, F., Valle, M., Martos, R., Ruz, F. J., Rios, R., Montilla, P., and Canete, R. Sex hormone-binding globulin as a marker for hyperinsulinemia and/or insulin resistance in obese children. Eur J Endocrinol 143, 85-9. (2000).
	7506666CD1	585825|Shbg	1.5E-102	[Mus musculus][Small molecule-binding protein][Extracellular (excluding cell wall)] Sex hormone binding globulin (androgen binding protein), binds and transports androgens and is involved in spermatogenesis; plasma levels of human SHBG are associated with obesity, heart disease, and noninsulin-dependent diabetes. Wang, Y. M., Sullivan, P. M., Petrusz, P., Yarbrough, W., and Joseph, D. R. The androgen-binding protein gene is expressed in CD1 mouse testis. Mol Cell Endocrinol 63, 85-92 (1989).
40	7511731CD1	g1911490	9.8E-144	[Homo sapiens] Con1
				Azen, E. A. et al., PRB1, PRB2, and PRB4 coded polymorphisms among human salivary concanavalin-A binding, II-1, and Po proline-rich proteins, Am. J. Hum. Genet. 58, 143-153 (1996).
		610704| PRB2	7.9E-145	[Homo sapiens][Extracellular (excluding cell wall)] Basic proline-rich salivary protein, a member of the BstNI-type subfamily of the salivary proline-rich protein (PRP) family.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Azen, E. A. et al., PRB2/1 fusion gene: a product of unequal and homologous crossing-over between proline-rich protein (PRP) genes PRB1 and PRB2., Am J Hum Genet 50, 842-51. (1992).
		626566| Prp	2.4E-65	[Mus musculus][Extracellular (excluding cell wall)] Proline rich protein with tandem repeats, expression is induced in salivary glands by isoproterenol and feeding tannins.
				Layfield, R. et al., cDNA clones for mouse parotid proline-rich proteins. mRNA regulation by isoprenaline and the nucleotide sequence of proline-rich protein cDNA MP5., Eur J Biochem 204, 591-7 (1992).
41	7511735CD1	g1911490	4.3E-111	[Homo sapiens] Con1
				Azen, E. A. et al. (supra)
		610704| PRB2	3.5E-112	[Homo sapiens][Extracellular (excluding cell wall)] Basic proline-rich salivary protein, a member of the BstNI-type subfamily of the salivary proline-rich protein (PRP) family.
				Azen, E. A. et al. (supra)
		329524| Rn.9842	3.8E-58	[Rattus norvegicus] Salivary proline-rich protein 15, mediates interferon gamma (Ifng) induced apoptosis.
				Deiss, L. P. et al., Identification of a novel serine/threonine kinase and a novel 15-kD protein as potential mediators of the gamma interferon-induced cell death., Genes And Development 9, 15-30 (1995).
42	7511729CD1	714990| ZK418.5	1.8E-28	[Caenorhabditis elegans] Putative transmembrane protein with similarity to human NIFIE14, putative seven transmembrane domain protein.
				Jiang, M. et al., Genome-wide analysis of developmental and sex-regulated gene expression profiles in Caenorhabditis elegans., Proc Natl Acad Sci U S A 98, 218-223 (2001).
43	7511255CD1	598940| FLJ11198	1.1E-49	[Homo sapiens][ATP-binding cassette; Active transporter, primary; Hydrolase; Transporter; ATPase] Member of the ABC transporter family.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
45	7511657CD1	334646| CHGA	5.8E-11	[Homo sapiens][Cytoplasmic; Extracellular (excluding cell wall)] Chromogranin A, member of a family of acidic proteins secreted with hormones and neurotransmitters; overexpression is associated with genetic hypertension.
				Taylor, C. V., Taupenot, L., Mahata, S. K., Mahata, M., Wu, H., Yasothornsrikul, S., Toneff, T., Caporale, C., Jiang, Q., Parmer, R. J., Hook, V.Y., and O'Connor, D. T. Formation of the catecholamine release-inhibitory peptide catestatin from chromogranin A. Determination of proteolytic cleavage sites in hormone storage granules. J Biol Chem 275, 22905-15 (2000).
49	7512332CD1	g190484	1.3E-77	[Homo sapiens] prepro salivary proline-rich protein.
				Maeda, N. et al., Differential RNA splicing and post-translational cleavages in the human salivary proline-rich protein gene system, J. Biol. Chem. 260, 11123-11130 (1985).
		610734| PRH1	1.1E-78	[Homo sapiens][Extracellular (excluding cell wall)] Proline-rich protein HaeIII subfamily 1, a member of the HaeIII-type subfamily of the proline rich protein (PRP) family, an acidic proline-rich salivary protein, contains a secretory signal sequence.
				Kim, H. S., and Maeda, N. Structures of two HaeIII-type genes in the human salivary proline-rich protein multigene family. J Biol Chem 261, 6712-8 (1986).
50	7511219CD1	g3719361	1.6E-17	[Homo sapiens] small inducible cytokine A23 precursor.
				Nomiyama, H. et al., Organization of the chemokine gene cluster on human chromosome 17q11.2 containing the genes for CC chemokine MPIF-1, HCC-2, HCC-1, LEC, and RANTES, J. Interferon Cytokine Res. 19, 227-234 (1999).
		341854| SCYA23	1.3E-18	[Homo sapiens][Ligand][Extracellular (excluding cell wall)] Small inducible cytokine A23, a CC chemokine that binds CC chemokine receptor 1 (CCR1) and induces calcium flux, inhibits proliferation of myeloid progenitor cells and induces chemotaxis in monocytes and resting T lymphocytes.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Ragno, S., Romano, M., Howell, S., Pappin, D. J., Jenner, P. J., and Colston, M. J. Changes in gene expression in macrophages infected with Mycobacterium tuberculosis: a combined transcriptomic and proteomic approach. Immunology 104, 99-108. (2001).
51	7510933CD1	598940| FLJ11198	4.9E-58	[Homo sapiens][ATP-binding cassette; Active transporter, primary; Hydrolase; Transporter; ATPase] Member of the ABC transporter family.
57	7512320CD1	731915| FLJ14454	2.9E-60	[Homo sapiens] Member of the secretin family of G protein-coupled receptors (GPCR), contains a latrophilin-CL-1-like G protein signaling (GPS) domain, which may bind ligands involved in cellular signaling, has weak similarity to leukocyte activation antigen (human CD97).
59	7512442CD1	g22859213	1.00E-162	MUC15/S protein [Homo sapiens]
60	7512311CD1	g10862826	5.5E-195	[Homo sapiens] (AJ299403) ADIR1
				Dron, M. et al., Molecular Cloning of ADIR, a Novel Interferon Responsive Gene Encoding a Protein Related to the Torsins., Genomics 79, 315-325 (2002).
		743758| TOR1B	3.5E-50	[Homo sapiens][Chaperones; Hydrolase; Heat shock protein; ATPase] Torsin family 1 member B, a member of the torsin gene family that has functional domains similar to those of the AAA; HSP; Clp-ATPase superfamily of chaperone-like proteins, expressed in the central nervous system.
				Ozelius, L. J. et al., The early-onset torsion dystonia gene (DYT1) encodes an ATP-binding protein., Nat Genet 17, 40-8 (1997).
				Kustedjo, K. et al., Torsin A and its torsion dystonia-associated mutant forms are lumenal glycoproteins that exhibit distinct subcellular localizations., J Biol Chem 275, 27933-9 (2000).
61	7512474CD1	g5705912	3.6E-75	[Homo sapiens] pregnancy-specific beta 1 glycoprotein; PSG
	7512474CD1			Panzetta-Dutari, G. M. et al., Nucleotide sequence of a pregnancy-specific beta 1 glycoprotein gene family member. Identification of a functional promoter region and several putative regulatory sequences., Mol. Biol. Rep. 16, 255-262 (1992).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
61 - Cont'd		337290| PSG5	6.9E-75	[Homo sapiens][Structural protein][Extracellular (excluding cell wall)] Pregnancy specific beta-1-glycoprotein 5, a member of the pregnancy-specific glycoprotein and carcinoembryonic antigen families that is expressed in the fetal liver and placenta, may play a role in pregnancy
				Koritschoner, N. P. et al., Analyses of cis-acting and trans-acting elements that are crucial to sustain pregnancy-specific glycoprotein gene expression in different cell types., Eur J Biochem 236, 365-72 (1996).

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7510186CD1	268	S8 S108 S128 S216 T74 T137 T148 T181	N93	signal_cleavage: M1-G35	SPSCAN
					Signal Peptide: M1-G35	HMMER
					Immunoglobulin domain: G56-A119, G153-A211	HMMER_PFAM
					Immunoglobulin: P145-S228, P48-A136	HMMER_SMART
					Immunoglobulin C-2 Type: V54-G124, E151-S216	HMMER_SMART
					Ig superfamily from SCOP: F141-P231, F44-I133	HMMER_INCY
2	7510045CD1	205	S21 S89 S195 T52 Y53	N130 N140	signal_cleavage: M1-A23	SPSCAN
					Signal Peptide: M1-P19, M1-S21, M1-G22, M1-A23, M1-G25	HMMER
3	7510258CD1	95	S9 S14 S27 S64 S80		signal_cleavage: M1-G46	SPSCAN
					Cytosolic domain: M1-R29 Transmembrane domain: L30-V52 Non-cytosolic domain: V53-R95	TMHMMER
4	7510450CD1	33			signal_cleavage: M1-A21	SPSCAN
5	7504544CD1	46	S16		signal_cleavage: M1-A13	SPSCAN
6	7504804CD1	899	S140 S201 S207 S256 S471 S537 S542 S553 S728 S829 T101 T358 T543 T597 T719 T752 T895 Y168 Y365	N184 N222 N312 N464 N469 N595 N634 N685 N704 N743 N758	signal_cleavage: M1-A38	SPSCAN
					Signal Peptide: M1-A38	HMMER
					TPR Domain: F295-F328, P689-C722, W619-Q652, D655-E688	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Tetratricopeptide repeats: P689-C722, F295-F328, W619-Q652, D655-E688	HMMER_SMART
					SIMILARITY TO S. CEREVISIAE PROTEASE A INHIBITOR 3 PROTEASE PD185879: W21-E142, F197-E310, W500-L716	BLAST_PRODOM
					Leucine zipper pattern: L651-L672, L658-L679, L695-L716	MOTIFS
					Cell attachment sequence: R433-D435	MOTIFS
7	7510156CD1	110	S65	N58 N63 N80	signal_cleavage: M1-A33	SPSCAN
					Signal Peptide: M18-A33, M13-A33	HMMER
					Cytosolic domain: M1-Y19 Transmembrane domain: L20-V42 Non-cytosolic domain: G43-S110	TMHMMER
8	7510157CD1	35			signal_cleavage: M1-Q33	SPSCAN
					Signal Peptide: M13-Q33, M18-Q33	HMMER
					Cytosolic domain: Q33-L35 Transmembrane domain: I15-Q32 Non-cytosolic domain: M1-G14	TMHMMER
9	7510993CD1	49	T19		signal_cleavage: M1-G46	SPSCAN
10	7511149CD1	84			signal_cleavage: M1-A56	SPSCAN
					Signal Peptide: M1-S26	HMMER
					Leucine zipper pattern: L50-L71	MOTIFS
11	7511184CD1	242	S135 S149 S224 T82 T94 T177	N50 N92 N160	Signal Peptide: M1-G20, M2-G20, M2-G23, M2-V26, M2-G30, M1-S25, M2-G24, M2-S25	HMMER
12	7511240CD1	33	T16 T20	N2	signal_cleavage: M1-R18	SPSCAN
					Signal Peptide: M1-T20	HMMER
13	7511376CD1	37	S20 S30 S31 T35		signal_cleavage: M1-S16	SPSCAN
					Signal Peptide: M1-S16, M1-S19, M1-S20, M1-K22, M1-T23, M1-S25	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	7501330CD1	65	T54 T58		signal_cleavage: M1-T54	SPSCAN
15	7509961CD1	149	S15 S88 S115 S117 T98 T113 T123		Cytosolic domain: G46-P149 Transmembrane domain: F23-L45 Non-cytosolic domain: M1-C22	TMHMMER
					PROTEIN ANCIENT UBIQUITOUS PRECURSOR AUP1 SIGNAL F44B9.5 CHROMOSOME III PD042305: E10-T113	BLAST_PRODOM
16	7509963CD1	267	S15 S88 S141 S148 S180 T98 T200 T251		signal_cleavage: M1-C37	SPSCAN
					Cytosolic domain: G46-A267 Transmembrane domain: F23-L45 Non-cytosolic domain: M1-C22	TMHMMER
					PROTEIN ANCIENT UBIQUITOUS PRECURSOR AUP1 SIGNAL F44B9.5 CHROMOSOME III PD042305: E10-V248	BLAST_PRODOM
17	7505089CD1	235	S42 S179 S209 S212 S223 T138 T145 Y74	N123	signal_cleavage: M1-R33	SPSCAN
					Cytosolic domain: K32-L235 Transmembrane domain: G10-V31 Non-cytosolic domain: M1-S9	TMHMMER
18	7510139CD1	286	S27 S66 S75 S93 S148 S212 S240 T99 T101 T208	N97	signal_cleavage: M1-S24	SPSCAN
					Signal Peptide: M1-S24, M7-A22	HMMER
19	7505053CD1	47	S44 T26		signal_cleavage: M1-T19	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
20	7511116CD1	240	S99 S101 S138 T48 T63 T73 T159	N79 N205 N223	signal_cleavage: M1-L29	SPSCAN
					Signal Peptide: M1-P24, M1-G25, M1-L29	HMMER
					EGF-like domain IPB000561: C181-G189	BLIMPS_BLOCKS
					PROTEIN GLYCOPROTEIN EGF-LIKE DOMAIN F09E8.2 HT PD031339: P33-G146	BLAST_PRODOM
					F09E8.2 PROTEIN GLYCOPROTEIN EGF-LIKE DOMAIN PD059364: F148-R204	BLAST_PRODOM
					EGF-like domain signature 1: C181-C192	MOTIFS
					EGF-like domain signature 2: C181-C192	MOTIFS
21	7511175CD1	193	S116 T114 T144 Y62		signal_cleavage: M1-P23	SPSCAN
					Signal Peptide: M1-P18, M1-P23, M1-G24, M1-G27	HMMER
					emp24/gp25L/p24 family: L13-A188	HMMER_PFAM
					Cytosolic domain: M1-K4 Transmembrane domain: I5-G27 Non-cytosolic domain: F28-A193	TMHMMER
					emp24/gp25L/p24 family. PF01105: G102-F108	BLIMPS_PFAM
22	7504660CD1	35			signal_cleavage: M1-S17	SPSCAN
					Signal Peptide: M1-V16, M1-L18, M1-T20, M1-A22	HMMER
					Cytosolic domain: R27-F35 Transmembrane domain: L4-W26 Non-cytosolic domain: M1-T3	TMHMMER
					Heat-stable enterotoxin IPB001489: M1-A15	BLIMPS_BLOCKS
					Connexin signature PR00206: K2-W26	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Adenosine A3 receptor signature PR00555: L4-M19	BLIMPS_PRINTS
					Rice seed allergenic protein RAG signature PR00809: L5-V16	BLIMPS_PRINTS
23	7504681CD1	77	S24		signal_cleavage: M1-G51	SPSCAN
					Cytosolic domain: M1-S30 Transmembrane domain: L31-G48 Non-cytosolic domain: P49-T77	TMHMMER
					Epithelial membrane prot PR01454: F34-W44	BLIMPS_PRINTS
24	7506472CD1	105		N25	signal_cleavage: M1-A17	SPSCAN
					Signal Peptide: M1-A17, M1-G19, M1-S20, M1-H22	HMMER
					Clusterin IPB000753: M1-A17	BLIMPS_BLOCKS
					Mammalian defensin IPB001271: M1-S20	BLIMPS_BLOCKS
					2Fe-2S ferredoxin signature PR00159: P84-C92	BLIMPS_PRINTS
					Glucagon family receptor PR01353: L5-Q18	BLIMPS_PRINTS
25	7506483CD1	30			signal_cleavage: M1-T25	SPSCAN
					Signal Peptide: M1-A24	HMMER
					Cytosolic domain: M1-N6 Transmembrane domain: W7-I29 Non-cytosolic domain: G30-G30	TMHMMER
					NAD-dependent glycerol-3-phosphate dehydrogenase signature PR00077: V13-G30, W7-I27	BLIMPS_PRINTS
					Neuropeptide Y2 receptor signature PR01014: G4-S21	BLIMPS_PRINTS
26	7506525CD1	170	S87 S154 T153		signal_cleavage: M1-G45	SPSCAN
					Signal Peptide: M1-S15, M1-A19, M1-G20, M1-P22	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Chitin-binding domain signature PR00451: A70-C78, G85-G92	BLIMPS_PRINTS
					Mollusc metallothionein signature PR00875: C91-L102, E90-G100	BLIMPS_PRINTS
					Neuropeptide Y6 receptor signature PR01017: A6-R21	BLIMPS_PRINTS
27	7506549CD1	978	S65 S133 S171 S198 S243 S264 S338 S459 S492 S500 S553 S604 S648 S649 S667 S743 S799 S950 S959 S964 T42 T209 T317 T331T346 T423 T482 T504 T506 T694 T836 T927 Y31	N196 N219 N262 N394 N421 N498 N513 N536 N551 N733	signal_cleavage: M1-C20	SPSCAN
					Signal Peptide: M1-A21, M1-K23, M1-C25, M1-G28, M1-C20, M1-G19	HMMER
					PKD domain: N622-V711, P528-V615, N716-V808, P433-V521	HMMER_PFAM
					Cadherin repeats.: L599-P719	HMMER_SMART
					Fibronectin type 3 domain: G593-S702, P434-T512, P719-S799, P305-G418	HMMER_SMART
					Delta-aminolevulinic acid dehydratase proteins BL00169: A642-V656	BLIMPS_BLOCKS
					Purothionin signature PR00288: G921-L935	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					CHITINASE A PRECURSOR HYDROLASE GLYCOSIDASE CHITIN DEGRADATION SIGNAL CHIA PD036147: P619-T801 PD036147: P528-T706	BLAST_PRODOM
28	7506683CD1	128	S46 S125 T35		signal_cleavage: M1-P23	SPSCAN
					Signal Peptide: M1-L28, M1-G33	HMMER
					Glutamine amidotransferases class-II proteins BL00443: P99-G108	BLIMPS_BLOCKS
					Inhibin beta B chain signature PR00671: G80-P99	BLIMPS_PRINTS
29	7510814CD1	31		N18 N23	signal_cleavage: M1-A19	SPSCAN
					Signal Peptide: M1-A19, M1-T21, M1-V22, M1-S24	HMMER
					Small CXC cytokine family signature PR00437: L9-Q17	BLIMPS_PRINTS
					Guanylin precursor signature PR00774: C10-V22	BLIMPS_PRINTS
30	7504727CD1	38			signal_cleavage: M1-C21	SPSCAN
					Signal Peptide: M1-T25	HMMER
					Pleiotrophin/midkine family signature PR00269: R6-Q22	BLIMPS_PRINTS
					Alpha-2C adrenergic receptor signature PR00560: L13-T25	BLIMPS_PRINTS
					Alanyl-tRNA synthetase signature PR00980: V9-Q22	BLIMPS_PRINTS
31	7506958CD1	32	S27		signal_cleavage: M1-D25	SPSCAN
32	7505332CD1	61	S8 T59		signal_cleavage: M1-A49	SPSCAN
					Kappa opioid receptor signature PR00532: V3-P18	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Type 3 bombesin receptor signature PR00637: C23-M40	BLIMPS_PRINTS
33	7505364CD1	41			signal_cleavage: M1-T19	SPSCAN
					Cytosolic domain: M1-I6 Transmembrane domain: F7-L29 Non-cytosolic domain: I30-C41	TMHMMER
					Cytidine and deoxycytidylate deaminases zinc-binding regions BL00903: E32-C41	BLIMPS_BLOCKS
					Glyceraldehyde 3-phosphate dehydrogenase active site: D3-T40	PROFILESCAN
					Hok/Gef cell toxic protein family signature PR00281: L8-L26	BLIMPS_PRINTS
					Mammalian lipoxygenase signature PR00467: F9-L26	BLIMPS_PRINTS
34	7505455CD1	54	S3		signal_cleavage: M1-T25	SPSCAN
					cGMP-dependent protein kinase signature PR00104: F14-G21	BLIMPS_PRINTS
					Alpha-1A adrenergic receptor signature PR00240: R16-A35	BLIMPS_PRINTS
					Beta-1 adrenergic receptor signature PR00561: G21-R40	BLIMPS_PRINTS
					Peptidase family U7. PF01343: S9-G22	BLIMPS_PFAM
35	7505641CD1	77	S22 S37 S57 T33		signal_cleavage: M1-S22	SPSCAN
					Signal Peptide: M1-S22	HMMER
					Lysozyme/alpha-lactalbumin superfamily signature PR00135: F56-V65	BLIMPS_PRINTS
					Histamine H1 receptor signature PR00530: F13-K24	BLIMPS_PRINTS
					Cholinesterase signature PR00878: G11-L19	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
36	7511242CD1	89	S45 S54		signal_cleavage: M1-T26	SPSCAN
					Signal Peptide: M1-A17, M1-L19	HMMER
					Cytosolic domain: Q23-K89 Transmembrane domain: I5-V22 Non-cytosolic domain: M1-E4	TMHMMER
					Neuropeptide Y2 receptor signature PR01014: V34-W49	BLIMPS_PRINTS
					HYDROLASE PROTEIN DIPEPT PD00794: G64-P73	BLIMPS_PRODOM
37	7506564CD1	229	S69 S108 S131 T22 T44 Y179	N29 N57 N67 N148	signal_cleavage: M1-A24	SPSCAN
					Signal Peptide: M1-A24, M1-A26, M1-G28, M8-A24, M8-A26	HMMER
					CUB domain: G38-Y143	HMMER_PFAM
					CUB domain first found in C1r, C1s, uEGF, and bone morphogenic protein 1 (BMP-1): C32-S146	HMMER_SMART
					CUB domain proteins profile BL01180: C85-G95	BLIMPS_BLOCKS
					Interleukin 8A receptor signature PR00572: V138-I151	BLIMPS_PRINTS
					Adrenomedullin signature PR00801: R6-A26	BLIMPS_PRINTS
38	7509076CD1	244	S3 S30 T73 T182	N50	signal_cleavage: M1-A26	SPSCAN
					Signal Peptide: M1-A26, M1-G25, M1-G28	HMMER
					EGF-like domain: C114-C141, C148-C183	HMMER_PFAM
					Epidermal growth factor-like domain.: I113-H142, E147-M184	HMMER_SMART
					Calcium-binding EGF-like domain: D144-M184	HMMER_SMART
					Calcium-binding EGF-like domain proteins pattern proteins BL01187: H142-T153, C159-L174	BLIMPS_BLOCKS
					Type III EGF-like signature PR00011: V123-C141	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					CALCIUM-BINDING PRECURSOR PD00919: C130-C159	BLIMPS_PRODOM
					PROTEIN GLYCOPROTEIN EGFLIKE DOMAIN PD031267: M1-I113 PD030385: G186-V234	BLAST_PRODOM
					EGF-like domain signature 1: C130-C141	MOTIFS
					EGF-like domain signature 2: C130-C141	MOTIFS
					Calcium-binding EGF-like domain pattern signature: D144-C168	MOTIFS
39	7506666CD1	293	S38 S71 S162 S198 S213 T9 T161 T231		signal_cleavage: M1-R25	SPSCAN
					Signal Peptide: M1-A29, M1-Q26,	HMMER
					Laminin G domain: V75-E205	HMMER_PFAM
					Laminin G domain: T67-K202	HMMER_SMART
					PROTEIN PRECURSOR SIGNAL SEX HORMONE-BINDING GLOBULIN SHBG STEROID-BINDING SBP TESTIS-SPECIFIC PD013322: Q26-E74	BLAST_PRODOM
					SEX HORMONE-BINDING GLOBULIN PRECURSOR SHBG STEROID-BINDING PROTEIN SBP TESTIS-SPECIFIC ANDROGEN-BINDING PD150870: E234-L288	BLAST_PRODOM
					SEX HORMONE-BINDING GLOBULIN DM01077|P04278|36-198: T36-W199, DM01077|P15196|1-163: T36-W199, DM01077|A39030|36-218: P35-D218 DM01077|P08689|36-199: P35-W199	BLAST_DOMO
40	7511731CD1	272	S24 S90 S152	N86 N128	signal_cleavage: M1-A16	SPSCAN
			S214		Signal Peptide: M1-A16, M1-N18, M1-A16, M1-N20	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
40 - Cont'd					PROTEIN REPEAT SIGNAL PRECURSOR PRION GLYCOPROTEIN NUCLEAR GPI-ANCHOR BRAIN MAJOR PD001091: P36-S268, K56-P270	BLAST_PRODOM
					PROTEIN REPEAT MICROTUBULE-ASSOCIATED MICROTUBULES PHOSPHORYLATION BASSOON ALTERNATIVE SPLICING LARGE PROLINE-RICH PD005493: L31-P261, P51-P270, Q78-Q272, P101-Q272	BLAST_PRODOM
					TRACHEAL COLONIZATION FACTOR PRECURSOR SIGNAL SARCALUMENIN CALCIUM-BINDING GLYCOPROTEIN ALTERNATIVE SPLICING PD136752: E21-G264, P29-P271, P51-Q272, S15-G230	BLAST_PRODOM
					PROLINE-RICH PROTEIN DM01281|P04280|126-210: G106-G188 G127-S214 G168-G253 G189-P271 G44-Q131 G65-S152 K138-Q228 K76-G167 Q25-S90 G34-G105 S24-G64	BLAST_DOMO
					PROLINE-RICH PROTEIN DM01281|P04280|17-124: G137-Q228 G75-G167 G96-G188 N107-G208 N169-P271 N45-G146 Q17-G126 S148-G257	BLAST_DOMO
					PROLINE-RICH PROTEIN DM01281|P04280|212-315: P112-G208 P135-Q245 P174-P270 P29-G141 P50-G146 P68-G183 R154-G257 R216-P271 R92-G203 S24-G121	BLAST_DOMO
					PROLINE-RICH PROTEIN DM01281|P10163|107-216: G106-S210 G127-Q248 G147-Q263 G168-P270 G189-P271 G34-Q131 G44-S148 G65-P174 N86-P195 S24-P112	BLAST_DOMO
41	7511735CD1	230	S24 S110 S172		signal_cleavage: M1-A16	SPSCAN
					Signal Peptide: M1-A16, M1-N18, M1-N20	HMMER
					PROTEIN REPEAT SIGNAL PRECURSOR PRION GLYCOPROTEIN NUCLEAR GPI-ANCHOR BRAIN MAJOR PD001091: P40-Q230, P36-S226, G38-P228, S14-P225	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					INTERMEDIATE CHAIN MULTIPLE BANDED ANTIGEN PRECURSOR SIGNAL PD103341: P41-S226, A33-Q186, P62-Q230, Q17-G152	BLAST_PRODOM
					PROTEIN REPEAT MICROTUBULE-ASSOCIATED MICROTUBULES PHOSPHORYLATION BASSOON ALTERNATIVE SPLICING LARGE PROLINE-RICH PD005493: S24-P219, L31-Q230, P74-S226	BLAST_PRODOM
					TRACHEAL COLONIZATION FACTOR PRECURSOR SIGNAL SARCALUMENIN CALCIUM-BINDING GLYCOPROTEIN ALTERNATIVE SPLICING PD136752: S14-G222, G34-Q230, P29-P229	BLAST_PRODOM
					PROLINE-RICH PROTEIN DM01281|P04280|17-124: G96-Q186 K107-G215 N127-P229 N66-G166 Q17-G146	BLAST_DOMO
					PROLINE-RICH PROTEIN DM01281|P04280|212-315: P111-G215 P132-P228 P29-G141 P50-G161 P71-G166 P89-Q203 R174-P229 S24-G120	BLAST_DOMO
					PROLINE-RICH PROTEIN DM01281|P04281|3-95: G157-P229 K107-P194 L19-R112 N127-R218 N45-P153 N66-R174 S168-P229	BLAST_DOMO
					PROLINE-RICH PROTEIN DM01281|P10163|107-216: G105-Q221 G126-P228 G147-P229 G34-Q151 G44-P153 G65-S168 G86-Q206 S24-R112	BLAST_DOMO
42	7511729CD1	126	S119 T20		signal_cleavage: M1-G62	SPSCAN
					DETHIOBIOTIN SYNTHETASE SYNTHASE. PD02561: W33-G39	BLIMPS_PRODOM
					Terpene synthase IPB002365: F6-Y29	BLIMPS_BLOCKS
					TRANSMEMBRANE SEVEN DOMAIN PROTEIN COSMID ZK418 PD155783: M1-I117	BLAST_PRODOM
43	7511255CD1	114	S36 S55 T3 T71		signal_cleavage: M1-F34	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
44	7511628CD1	40			signal_cleavage: M1-A32	SPSCAN
					Signal Peptide: P15-A32, P10-A32, G8-A32, M1-A32, P10-A32, P13-A32	HMMER
					Nerve growth factor family IPB002072: P15-D39	BLIMPS_BLOCKS
					Glial cell line-derived neurotrophic factor recept PR01319: P13-L25	BLIMPS_PRINTS
45	7511657CD1	38		N33	signal_cleavage: M1-A18	SPSCAN
					Signal Peptide: M1-V16, M1-A18, M1-P20, M1-S23	HMMER
46	7512343CD1	334	S76 S82 S294	N74 N87 N130	signal_cleavage: M1-G27	SPSCAN
			S328 T28 T44 T132 T197	N143 N160 N173 N263 N292	Signal Peptide: M1-R30, M1-G24, M1-G25, M1-G27	HMMER
					Cytosolic domain: C229-L334 Transmembrane domain: I206-V228Non-cytosolic domain: M1-G205	TMHMMER
					PROTEIN TRANSMEMBRANE INTERGENIC REGION RECQ-PLD. PD01736: G205-L217, G114-A125	BLIMPS_PRODOM
					GLYCOPROTEIN ANTIGEN PRECURSOR IMMUNOGLO. PD02327: L92-V103, A19-Q40	BLIMPS_PRODOM
					PROTEIN MEMBRANE TRANSPORT TRANS. PD02364: G211-V228	BLIMPS_PRODOM
47	7512357CD1	162	S39 S74 S76 S104		signal_cleavage: M1-Y18	SPSCAN
			S109 T27 T72 T99		Signal Peptide: M1-Y18	HMMER
48	7511046CD1	37	S10		signal_cleavage: M1-G30	SPSCAN
49	7512332CD1	187	S15 S24 S59		signal_cleavage: M1-A16	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					SALIVARY ACIDIC PROLINE-RICH PHOSPHOPROTEIN 1/2 PRECURSOR PRP1/PRP 3 PRP2/PRP4 PIFF/PIFS PROTEIN A/PROTEIN C CONTAINS: PEPTIDE PC REPEAT SALIVA SIGNAL PAROTID GLAND PHOSPHORYLATION P PD054888: M1-D67	BLAST_PRODOM
					PROTEIN REPEAT SIGNAL PRECURSOR PRION GLYCOPROTEIN NUCLEAR GPI-ANCHOR BRAIN MAJOR PD001091: R46-Q187	BLAST_PRODOM
					INTERMEDIATE CHAIN MULTIPLE BANDED ANTIGEN PRECURSOR SIGNAL PD103341: E39-P186, Q47-Q187	BLAST_PRODOM
					TRACHEAL COLONIZATION FACTOR PRECURSOR SIGNAL SARCALUMENIN CALCIUM-BINDING GLYCOPROTEIN ALTERNATIVE SPLICING PD136752: S33-G175, D34-Q187	BLAST_PRODOM
					PROLINE-RICH PROTEIN DM01281|P04280|17-124: G127-Q187 Q17-G136 Q80-G174	BLAST_DOMO
					PROLINE-RICH PROTEIN DM01281|P10163|107-216: E45-R143 G68-P159 G78-P180 Q134-P181	BLAST_DOMO
					PROLINE-RICH PROTEIN DM01369|P02810|86-164: G68-R145 P107-P186 P49-R128 P86-P160	BLAST_DOMO
					PROLINE-RICH PROTEIN DM03894|A39066|1-159: G36-Q182 G61-Q187	BLAST_DOMO
50	7511219CD1	67	S48		signal_cleavage: M1-G18	SPSCAN
					Signal Peptide: M1-G18, M1-A21, M1-V23, M1-D26, M1-G18	HMMER
					PRECURSOR CYTOKINE MACROPHAGE INFLAMMATORY PROTEIN MIP5 CHEMOKINE CC2 HCC2 NCC3 PD027876: M1-T46	BLAST_PRODOM
51	7510933CD1	150	S36 S55 S116 T3		signal_cleavage: M1-F34	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			T71		Cytosolic domain: M1-R124 Transmembrane domain: G125-I144 Non-cytosolic domain: S145-S150	TMHMMER
					Alpha/beta hydrolase fold signature PR00111: L136-F149	BLIMPS_PRINTS
52	7511461CD1	49	T7		signal_cleavage: M1-G41	SPSCAN
53	7511808CD1	30	T23		signal_cleavage: M1-A17	SPSCAN
					Signal Peptide: M1-A17	HMMER
					ATP synthase alpha and beta subunits signature: I5-S30	PROFILESCAN
54	7511817CD1	55			signal_cleavage: M1-G27	SPSCAN
					Signal Peptide: M1-A19	HMMER
					Signal Peptide: M1-F21	HMMER
					Thiamine pyrophosphate enzymes signature: A4-R54	PROFILESCAN
55	7511832CD1	36	S4		signal_cleavage: M1-G27	SPSCAN
					Signal Peptide: M1-T25	HMMER
					Signal Peptide: M1-G27	HMMER
					Signal Peptide: M1-K26	HMMER
					Cytosolic domain: M1-R6 Transmembrane domain: F7-G24 Non-cytosolic domain: T25-Q36	TMHMMER
56	7512301CD1	76	S23 S33		signal_cleavage: M1-S58	SPSCAN
57	7512320CD1	119	S3 T47	N82	signal_cleavage: M1-G26	SPSCAN
					Signal Peptide: M1-G22	HMMER
					Signal Peptide: M1-G26	HMMER
					EGF-like domain IPB000561: C63-G71	BLIMPS_BLOCKS
					Cytosolic domain: M1-V11 Transmembrane domain: L12-I34 Non-cytosolic domain: R35-T119	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					EGF-like domain signature 1: C63-C74	MOTIFS
58	7512371CD1	102	S3 S50		signal_cleavage: M1-A51	SPSCAN
59	7512442CD1	311	S16 S84 S118	N57 N88 N106	signal_cleavage: M1-G50	SPSCAN
			S126 T10 T27	N117 N175 N182	Signal Peptide: M28-G50	HMMER
			T184 T220 T240 T282	N190 N245 N252 N275 N284		
60	7512311CD1	350	S59 S107 S246	N122	signal_cleavage: M1-P21	SPSCAN
			S342 T124 T315		Signal Peptide: M1-S27	HMMER
			Y138		Signal Peptide: M1-G25	HMMER
					Signal Peptide: M1-P21	HMMER
					PROTEIN TORSIN A ATP-BINDING ENDOPLASMIC RETICULUM GLYCOPROTEIN: PD098108:T124-R344	BLAST_PRODOM
61	7512474CD1	157	T14 T96 T113	N104 N111	signal_cleavage: M1-A63	SPSCAN
					Signal Peptide: M1-A34	HMMER
					PRECURSOR GLYCOPROTEIN SIGNAL PREGNANCY-SPECIFIC CARCINOEMBRYONIC ANTIGEN IMMUNOGLOBULIN FOLD PREGNANCY MULTIGENE PD000677: N29-T113	BLAST_PRODOM
			T14 T96 T113	N104 N111	CARCINOEMBRYONIC ANTIGEN PRECURSOR AMINO-TERMINAL DOMAIN DM00372 |A54312|38-148: I38-Y143 |P11462|38-128: I38-D129 |JC4121|38-128: I38-D129 |A26902|38-146: I38-P144	BLAST_DOMO
					Cell attachment sequence: R127-D129	MOTIFS

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
62/7510186CB1 2781	1-2772, 103-686, 104-686, 564-1034, 564-1038, 564-1042, 729-1565, 869-1369, 1456-1686, 1687-1930, 1928-2178, 1928-2181, 1928-2282, 1928-2357, 1928-2454, 2000-2777, 2039-2537, 2063-2327, 2146-2777, 2208-2776, 2225-2775, 2229-2753, 2255-2470, 2264-2778, 2327-2762, 2328-2758, 2328-2762, 2330-2637, 2330-2669, 2330-2718, 2359-2768, 2384-2781, 2427-2762, 2503-2768
63/7510045CB1 1371	1-230, 1-1158, 478-1286, 494-1317, 747-1306, 796-1017, 799-1094, 800-992, 810-1093, 852-1371, 853-1104, 853-1152, 854-1136, 862-1139, 878-1136, 886-1158, 901-1132, 903-1135, 910-1175, 915-1318, 923-1007, 928-1153, 930-1020, 930-1084, 930-1090, 930-1103, 930-1105, 930-1108, 930-1117, 930-1134, 930-1167, 930-1179, 930-1269, 930-1298, 931-1032, 931-1055, 931-1074, 931-1078, 931-1094, 931-1095, 931-1112, 931-1133, 931-1164, 931-1176, 934-1077, 944-1098, 968-1152, 970-1114, 972-1177, 1051-1351
64/7510258CB1 2531	1-584, 1-2526, 11-184, 12-584, 14-584, 49-2505, 86-584, 193-584, 1200-1537, 1200-1551, 1200-1969, 1200-1990, 1205-1964, 1206-1839, 1364-2132, 1648-2404, 1661-2523, 1694-2404, 1843-2523, 1929-2531, 2128-2526
65/7510450CB1 631	1-630, 1-631
66/7504544CB1 1019	1-140, 1-935, 15-158, 20-158, 23-158, 24-279, 28-158, 28-335, 43-102, 44-102, 44-321, 45-128, 45-154, 45-282, 45-304, 45-366, 45-562, 46-158, 46-335, 51-341, 56-158, 59-284, 64-158, 157-288, 157-423, 157-447, 157-838, 169-314, 169-375, 169-392, 171-416, 171-461, 179-427, 181-666, 181-693, 184-502, 188-817, 189-594, 193-463, 194-483, 200-693, 203-526, 210-666, 211-775, 211-781, 214-411, 217-316, 217-466, 217-479, 220-506, 221-481, 222-491, 226-462, 226-478, 231-357, 231-461, 231-497, 231-521, 232-665, 233-517, 237-666, 240-664, 241-774, 253-526, 261-531, 261-634, 266-468, 266-486, 266-525, 266-538, 266-602, 266-934, 268-570, 271-740, 274-540, 274-784, 279-484, 279-562, 283-531, 287-579, 288-534, 288-555, 288-561, 292-567, 294-569, 295-709, 295-751, 296-582, 296-626, 297-465, 297-740, 298-576, 300-494, 300-530, 301-575, 309-580, 315-480, 315-935, 323-591, 323-610, 328-751, 333-816, 339-713, 340-609, 345-548, 345-886, 346-646, 349-634, 350-623, 351-540, 354-651, 356-835, 356-853, 367-615, 369-647, 370-673,
	371-613, 372-617, 372-621, 379-642, 383-560, 389-610, 389-654, 390-574, 390-676, 390-693, 391-647, 396-658, 396-660, 396-687, 404-842, 407-673, 408-856, 410-700, 411-530, 411-959, 420-705, 424-950, 427-766, 429-1019, 435-847, 443-675, 443-726, 443-922, 446-731, 446-969, 447-721, 451-817, 456-952, 457-970, 461-734, 465-686, 487-790, 489-739, 497-756, 505-746, 510-767, 511-680, 515-789, 517-737, 528-924, 530-780, 541-792, 543-784, 544-762, 544-849, 546-810, 554-897, 555-841, 556-830, 563-842, 564-774, 564-781, 565-844, 568-839, 583-823, 589-852, 593-859, 594-1007, 596-822, 596-857, 596-866, 605-829, 605-837, 605-855, 609-916, 627-745, 627-847, 627-887, 627-898, 672-900, 683-924, 687-933, 688-930, 702-823, 717-926, 726-928

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
67/7504804CB1 3590	1-253, 1-3590, 14-272, 14-374, 21-773, 26-663, 36-585, 38-402, 38-782, 39-351, 39-782, 41-293, 42-294, 44-308, 44-714, 48-307, 48-649, 50-598, 57-526, 98-422, 145-759, 169-631, 172-801, 190-836, 250-505, 277-484, 281-640, 379-1050, 386-774, 394-1087, 400-615, 416-684, 445-756, 483-1148, 484-769, 503-1151, 507-978, 519-681, 519-772, 595-775, 602-877, 609-838, 619-1398, 648-1061, 665-1267, 795-1432, 812-1145, 833-1385, 836-1489, 863-1114, 962-1273, 994-1260, 1072-1657, 1100-1842, 1113-1330, 1113-1513, 1143-1352, 1193-1440, 1374-1637, 1450-1660, 1454-1689, 1454-1916, 1454-1926, 1471-1547, 1473-2074, 1473-2165, 1473-2275, 1485-1675, 1523-1757, 1541-2009, 1542-2009, 1555-2027, 1571-2036, 1576-2179, 1586-1929, 1591-1827, 1613-1893, 1628-1967, 1678-1891, 1685-2181, 1703-2043, 1708-2289, 1711-2289, 1712-1966, 1738-1978, 1760-2289, 1761-1944, 1761-2272, 1787-2286, 1792-2190, 1815-1940, 1837-2112, 1838-2026, 1891-2117, 1909-2008, 1916-2228, 1924-2166, 1933-2289, 1959-2195, 1959-2210, 1985-2289, 2059-2289, 2141-2289,
	2209-2630, 2287-2677, 2287-2776, 2292-2569, 2302-2701, 2303-2776, 2312-2589, 2314-2629, 2323-2698, 2334-2566, 2334-2964, 2335-2612, 2336-2698, 2355-2568, 2387-2608, 2399-2631, 2402-2644, 2419-2529, 2434-2696, 2457-2700, 2468-2889, 2469-2730, 2471-2698, 2482-2943, 2505-3047, 2516-2695, 2525-2972, 2535-2904, 2535-3016, 2566-2721, 2572-2998, 2573-2994, 2582-3040, 2594-3040, 2623-2925, 2668-2836, 2675-3047, 2689-2945, 2700-2963, 2759-2971, 2785-3015, 2790-3017, 2797-3040, 2887-3036, 2887-3037, 2928-3590, 3034-3058, 3034-3075, 3034-3076, 3034-3085, 3034-3092, 3034-3123, 3034-3161, 3034-3166, 3034-3214, 3034-3215, 3034-3228, 3034-3230, 3034-3239, 3034-3244, 3034-3259, 3034-3277, 3034-3321, 3034-3335, 3034-3338, 3034-3342, 3034-3348, 3034-3360, 3034-3390, 3034-3409, 3108-3361, 3190-3431, 3190-3437, 3258-3338
68/7510156CB1 2365	1-223, 1-224, 1-323, 1-779, 1-806, 1-2046, 6-224, 20-284, 27-224, 51-224, 99-316, 477-893, 705-1296, 759-1065, 937-1531, 1113-1877, 1113-1878, 1114-1877, 1114-1878, 1122-1876, 1123-1878, 1164-1878, 1182-1297, 1182-1433, 1191-1804, 1277-1878, 1366-2005, 1849-2365
69/7510157CB1 2548	1-726, 10-2229, 11-691, 14-906, 243-570, 888-1479, 942-1248, 1120-1714, 1296-2061, 1305-2059, 1305-2060, 1305-2061, 1322-2061, 1347-2061, 1365-1480, 1365-1616, 1374-1908, 1460-2061, 1549-2092, 2032-2548
70/7510993CB1 1584	1-232, 2-1563, 3-280, 16-254, 16-260, 16-268, 17-253, 17-273, 21-277, 21-279, 23-173, 26-231, 27-261, 36-133, 36-196, 37-298, 38-262, 39-170, 39-242, 39-256, 39-270, 39-283, 40-275, 42-216, 42-273, 42-281, 42-309, 42-311, 42-312, 42-325, 43-238, 45-316, 46-295, 46-296, 46-311, 46-321, 48-300, 48-303, 49-294, 49-315, 50-236, 50-260, 50-297, 50-303, 50-318, 53-296, 56-296, 56-309, 58-265, 58-297, 65-292, 65-311, 86-342, 93-282, 115-389, 125-285, 195-261, 195-363, 296-547, 296-557, 352-627, 391-968, 414-711, 419-1085, 426-1046, 426-1056, 426-1109, 430-952, 440-736, 442-1329, 445-708, 446-1023, 448-741, 449-718, 455-744, 457-920, 471-978, 476-698, 476-715, 476-773, 477-788, 477-1109, 484-808, 484-918, 485-1064, 486-671, 487-725, 498-764, 507-770, 508-983, 509-682, 515-1060, 524-1054, 525-755, 525-1087, 530-1111, 535-767, 537-794, 540-713, 540-776, 540-1159, 542-1188, 544-828, 549-708, 550-780, 555-879, 560-1184, 568-783, 568-804, 568-835, 571-1126, 577-860, 577-1164, 580-820, 583-878, 583-1269, 585-801,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	585-1192, 588-884, 588-885, 592-1139, 592-1188, 596-835, 597-839, 600-1016, 607-901, 607-1124, 609-863, 611-1242, 617-878, 623-1096, 624-897, 625-868, 626-969, 628-878, 628-1278, 629-800, 632-1306, 636-1207, 640-746, 640-846, 640-866, 640-909, 646-1232, 646-1283, 647-1232, 652-1227, 652-1231, 653-1277, 655-921, 655-948, 657-992, 664-948, 667-859, 669-814, 671-1331, 672-859, 673-951, 675-948, 675-1231, 677-950, 678-1057, 687-897, 692-1199, 692-1228, 694-914, 694-920, 694-1127, 694-1347, 695-913, 695-947, 709-1283, 714-1443, 722-991, 724-1545, 732-984, 737-884, 738-1437, 739-1012, 739-1390, 744-1017, 751-1313, 754-1015, 754-1023, 757-1012, 766-902, 766-1031, 766-1307, 772-900, 773-1013, 773-1037, 773-1242, 774-972, 777-1008, 778-1030, 778-1054, 778-1400, 779-1328, 780-1229, 780-1239, 780-1536, 783-1056, 787-1254, 796-1000, 797-1010, 798-1339, 800-1077, 800-1257, 805-1103, 805-1322, 806-1444, 808-941, 812-1046, 817-1431, 821-1362, 821-1545, 825-1039, 825-1107, 826-1105, 834-1115, 835-1018, 837-1022, 837-1098,
	837-1132, 838-1451, 838-1511, 840-1032, 840-1049, 840-1083, 840-1097, 841-1089, 845-1097, 847-1378, 851-1156, 851-1502, 857-1303, 859-958, 860-1098, 860-1105, 860-1120, 861-1432, 862-1010, 863-1115, 865-1083, 866-1130, 866-1288, 866-1517, 876-1120, 877-1153, 877-1549, 877-1557, 886-1395, 887-1118, 887-1508, 892-1463, 895-1146, 898-1181, 899-1220, 905-1441, 907-1213, 913-1451, 914-1240, 915-1183, 916-1546, 921-1166, 924-1182, 924-1210, 924-1495, 925-1128, 926-1435, 932-1209, 942-1149, 946-1188, 950-1522, 951-1463, 956-1245, 956-1540, 960-1240, 961-1249, 963-1172, 964-1535, 970-1217, 970-1520, 976-1537, 978-1463, 980-1289, 984-1279, 990-1128, 992-1565, 993-1530, 996-1478, 998-1495, 999-1562, 999-1564, 1000-1222, 1000-1247, 1000-1560, 1001-1236, 1001-1567, 1005-1278, 1006-1463, 1007-1261, 1014-1308, 1016-1232, 1017-1230, 1017-1283, 1021-1259, 1021-1269, 1026-1223, 1030-1557, 1032-1083, 1032-1325, 1032-1467, 1048-1548, 1049-1444, 1050-1367, 1052-1548, 1053-1297, 1054-1566, 1055-1549, 1055-1570, 1057-1548,
	1058-1549, 1059-1344, 1059-1557, 1060-1463, 1062-1419, 1065-1283, 1065-1298, 1066-1310, 1066-1417, 1074-1312, 1075-1360, 1076-1570, 1078-1342, 1078-1567, 1081-1548, 1083-1549, 1085-1548, 1089-1546, 1089-1584, 1090-1548, 1090-1566, 1091-1548, 1093-1557, 1095-1548, 1096-1549, 1098-1549, 1102-1548, 1102-1566, 1103-1547, 1103-1549, 1103-1566, 1104-1557, 1105-1567, 1106-1549, 1106-1568, 1107-1567, 1109-1411, 1111-1565, 1115-1567, 1117-1549, 1121-1568, 1123-1567, 1127-1549, 1127-1570, 1128-1360, 1128-1548, 1129-1549, 1129-1564, 1129-1570, 1131-1566, 1132-1566, 1134-1400, 1135-1556, 1136-1474, 1139-1568, 1140-1574, 1150-1410, 1150-1418, 1153-1378, 1157-1568, 1158-1429, 1162-1549, 1168-1566, 1172-1440, 1173-1452, 1175-1412, 1175-1566, 1175-1568, 1175-1574, 1184-1426, 1197-1536, 1202-1548, 1208-1567, 1211-1451, 1213-1449, 1213-1459, 1213-1548, 1215-1455, 1218-1486, 1224-1448, 1226-1370, 1226-1464, 1226-1482, 1226-1495, 1230-1513, 1242-1358, 1245-1447, 1246-1496, 1258-1485, 1292-1510, 1307-1548, 1322-1570, 1359-1563, 1371-1460, 1412-1539, 1420-1549, 1430-1556, 1471-1549, 1478-1581, 1485-1557, 1486-1549
71/7511149CB1 1982	1-198, 1-363, 1-740, 1-1982, 37-636, 38-636, 140-511, 197-810, 381-968, 423-1010, 598-1166, 600-1149, 690-1263, 690-1281, 702-1267, 722-1222, 880-1503, 993-1513, 1000-1557, 1060-1617, 1081-1716, 1399-1982, 1563-1879, 1870-1982

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
72/7511184CB1 3246	1-238, 1-3246, 4-456, 86-749, 125-782, 286-925, 312-848, 405-1174, 550-1168, 602-1289, 612-1152, 612-1225, 612-1250, 617-1271, 709-1277, 745-1380, 762-1330, 803-1422, 868-1650, 940-1143, 949-1632, 1049-1726, 1086-1682, 1113-1715, 1116-1725, 1134-1589, 1141-1685, 1206-1888, 1213-1743, 1221-1733, 1250-2048, 1251-1865, 1256-1883, 1277-1848, 1290-1893, 1297-1847, 1312-1914, 1314-1851, 1318-1952, 1325-1977, 1339-2086, 1350-2023, 1376-2195, 1417-2075, 1423-2029, 1542-2201, 1553-2223, 1566-2049, 1599-2427, 1650-2242, 1681-2239, 1695-2239, 1784-2439, 1950-2465, 1952-2144, 2044-2721, 2065-2735, 2199-2737, 2393-2924, 2697-3246, 2713-3194, 3150-3246
73/7511240CB1 2012	1-229, 1-268, 1-366, 6-229, 6-2012, 19-228, 23-83, 23-90, 40-630, 227-473, 227-499, 227-581, 227-720, 227-784, 227-871, 227-923, 252-831, 265-970, 294-602, 298-836, 303-903, 501-888, 527-1048, 575-888, 705-1328, 734-960, 769-1363, 851-1427, 861-1119, 920-1577, 935-1550, 942-1528, 990-1477, 1037-1263, 1056-1945, 1058-1729, 1075-1360, 1076-1314, 1183-1474, 1208-1514, 1306-1724, 1336-1813, 1342-1673, 1342-1677, 1481-1768, 1582-1834, 1598-1890, 1599-1873, 1645-1928, 1657-1918, 1709-2009
74/7511376CB1 855	1-441, 1-840, 113-217, 113-319, 113-623, 135-392, 135-726, 140-421, 161-425, 163-729, 168-615, 188-366, 212-742, 215-479, 235-529, 238-480, 241-487, 248-809, 259-396, 262-660, 262-819, 267-620, 268-599, 270-627, 271-550, 287-532, 287-541, 313-400, 313-527, 328-502, 339-556, 340-617, 340-655, 348-488, 380-643, 545-735, 584-840, 585-855, 586-784
75/7501330CB1 1402	1-255, 1-1399, 19-220, 19-271, 32-246, 32-247, 33-274, 55-352, 55-354, 179-429, 197-452, 202-842, 204-437, 204-449, 211-493, 250-495, 252-790, 252-835, 261-490, 264-513, 264-833, 284-550, 294-417, 328-829, 346-950, 350-918, 352-881, 354-910, 357-680, 381-527, 395-972, 398-666, 417-640, 417-862, 417-956, 417-968, 426-671, 458-1049, 493-1036, 493-1089, 498-1070, 499-1193, 507-755, 510-1036, 512-773, 512-1068, 512-1098, 524-1351, 525-758, 531-1011, 540-925, 541-1006, 551-1099, 554-1134, 561-875, 564-841, 564-1037, 582-858, 582-859, 582-881, 588-1036, 591-1277, 594-911, 636-860, 641-1192, 641-1311, 654-935, 660-1037, 663-1193, 665-883, 665-960, 665-1397, 665-1398, 665-1402, 667-1161, 668-1126, 668-1343, 668-1371, 672-1371, 673-1402, 676-930, 682-1037, 691-1402, 692-1203, 697-1384, 699-972, 699-1037, 701-1149, 707-1384, 711-1384, 713-1384, 715-1128, 718-972, 718-1037, 718-1099, 722-1158, 732-1384, 735-1362, 737-1387, 739-1016, 739-1371, 747-972, 747-1037, 747-1158, 747-1190, 747-1193, 748-1099, 748-1193,
	748-1236, 756-1395, 760-1371, 765-1268, 769-1171, 774-972, 774-1036, 774-1037, 774-1116, 774-1158, 774-1242, 774-1247, 775-1242, 782-1177, 786-1371, 786-1399, 794-1039, 794-1396, 809-1387, 821-1402, 837-1398, 843-1391, 849-1090, 854-1149, 854-1158, 855-1181, 864-1370, 865-1095, 874-1399, 886-1402, 895-1141, 911-1329, 915-1381, 916-1148, 916-1153, 922-1263, 923-1399, 925-1124, 938-1121, 938-1401, 943-1194, 945-1386, 947-1235, 950-1121, 951-1243, 951-1389, 973-1385, 988-1385, 995-1190, 999-1389, 1011-1205, 1011-1385, 1012-1275, 1012-1386, 1017-1295, 1029-1228, 1032-1346, 1032-1385, 1038-1254, 1038-1383, 1040-1307, 1040-1329, 1040-1385, 1042-1385, 1056-1284, 1106-1387, 1117-1402, 1130-1389, 1182-1306, 1243-1385

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
76/7509961CB1 1697	1-185, 1-265, 1-1697, 61-665, 61-679, 64-408, 65-214, 65-335, 65-420, 65-485, 65-488, 65-641, 65-700, 74-563, 77-615, 82-557, 83-684, 87-594, 88-399, 89-649, 102-524, 147-415, 178-441, 188-453, 236-504, 243-788, 258-697, 267-779, 299-519, 299-568, 318-946, 363-963, 383-845, 386-618, 387-615, 389-862, 401-825, 404-679, 413-684, 415-888, 433-944, 443-735, 451-734, 454-854, 455-1012, 457-826, 457-833, 487-1026, 488-1067, 490-956, 502-625, 526-1085, 551-1080, 566-791, 567-823, 571-811, 579-1082, 589-848, 589-854, 599-924, 602-1188, 603-1107, 603-1229, 614-1214, 618-1215, 618-1277, 619-1077, 624-849, 625-974, 630-1127, 632-1230, 635-966, 638-879, 638-1318, 644-908, 645-948, 645-1165, 646-895, 646-904, 646-942, 647-877, 647-913, 648-887, 649-803, 655-889, 657-881, 665-1276, 666-971, 669-922, 669-976, 669-1230, 681-938, 681-946, 681-974, 681-1118, 688-982, 688-987, 690-1078, 690-1317, 699-940, 704-1345, 705-956, 710-982, 711-961, 713-987, 714-969, 721-1210, 722-993, 723-1013, 723-1265, 733-1362, 734-1197, 743-1242,
	746-1231, 748-1341, 753-1020, 755-829, 756-1277, 757-1008, 759-1085, 759-1688, 767-1122, 767-1514, 777-1143, 787-1688, 788-1236, 791-1062, 791-1682, 792-1054, 793-1196, 798-1441, 801-1405, 809-1415, 813-1688, 817-1688, 818-1011, 823-1004, 825-1299, 828-1278, 831-1097, 832-1079, 832-1265, 834-1688, 839-1310, 839-1454, 840-1141, 840-1148, 841-1084, 845-1099, 846-1051, 847-1063, 850-1692, 851-1014, 851-1140, 856-1404, 861-1158, 865-1023, 865-1058, 865-1109, 865-1146, 868-1315, 869-1001, 869-1115, 869-1405, 877-1120, 879-1096, 880-1147, 882-1109, 882-1242, 883-1177, 887-1169, 893-1139, 893-1691, 895-1514, 896-1688, 902-1428, 903-1384, 907-1108, 919-1196, 925-1178, 925-1688, 926-1163, 926-1179, 930-1203, 932-1069, 932-1092, 932-1235, 932-1492, 932-1525, 934-1100, 946-1691, 956-1479, 957-1345, 959-1495, 960-1152, 961-1674, 962-1565, 966-1190, 966-1201, 968-1480, 969-1242, 972-1510, 973-1233, 974-1243, 975-1268, 987-1229, 992-1194, 994-1238, 995-1609, 998-1256, 998-1656, 999-1267, 1000-1239, 1000-1264, 1000-1435,
	1000-1445, 1002-1656, 1013-1201, 1015-1114, 1015-1269, 1015-1277, 1017-1473, 1023-1229, 1024-1697, 1025-1167, 1025-1345, 1026-1348, 1027-1465, 1029-1500, 1030-1530, 1035-1622, 1038-1331, 1040-1237, 1040-1264, 1042-1265, 1043-1265, 1046-1314, 1047-1194, 1048-1308, 1053-1373, 1056-1293, 1058-1280, 1058-1318, 1064-1671, 1067-1292, 1068-1688, 1069-1350, 1069-1407, 1070-1623, 1071-1323, 1071-1337, 1073-1185, 1073-1657, 1074-1590, 1075-1248, 1079-1184, 1079-1664, 1082-1311, 1088-1299, 1090-1642, 1098-1351, 1101-1332, 1110-1572, 1112-1394, 1112-1596, 1116-1376, 1116-1648, 1116-1690, 1117-1688, 1119-1420, 1123-1368, 1123-1376, 1123-1378, 1123-1651, 1129-1327, 1129-1690, 1130-1657, 1138-1353, 1139-1384, 1143-1670, 1145-1697, 1146-1579, 1148-1477, 1149-1430, 1149-1441, 1149-1697, 1154-1652, 1155-1413, 1156-1439, 1160-1682, 1163-1578, 1163-1697, 1165-1436, 1170-1677, 1174-1463, 1178-1514, 1183-1682, 1183-1685, 1187-1441, 1187-1456, 1187-1646, 1187-1685, 1193-1688, 1194-1686, 1196-1476, 1200-1430, 1201-1490, 1212-1688,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1213-1697, 1215-1690, 1216-1688, 1220-1678, 1224-1692, 1226-1690, 1228-1488, 1228-1688, 1229-1689, 1229-1692, 1230-1687, 1231-1688, 1233-1688, 1234-1688, 1235-1487, 1235-1499, 1236-1683, 1238-1688, 1240-1470, 1240-1474, 1240-1644, 1242-1494, 1242-1690, 1247-1688, 1248-1658, 1248-1697, 1250-1496, 1252-1533, 1252-1550, 1255-1489, 1255-1523, 1255-1645, 1255-1692, 1255-1697, 1256-1556, 1256-1690, 1258-1697, 1259-1688, 1260-1697, 1261-1697, 1263-1697, 1264-1689, 1266-1548, 1269-1445, 1269-1556, 1269-1688, 1269-1690, 1270-1469, 1270-1691, 1272-1688, 1274-1652, 1276-1688, 1279-1611, 1280-1688, 1281-1686, 1281-1688, 1284-1692, 1285-1559, 1286-1692, 1287-1690, 1288-1685, 1289-1459, 1290-1692, 1292-1569, 1292-1690, 1292-1697, 1294-1690, 1295-1690, 1296-1688, 1297-1697, 1299-1529, 1299-1690, 1300-1688, 1303-1643, 1303-1683, 1303-1688, 1303-1690, 1307-1558, 1308-1688, 1309-1571, 1311-1688, 1313-1688, 1314-1687, 1315-1688, 1315-1689, 1319-1689, 1321-1625, 1321-1690, 1322-1697, 1324-1690, 1325-1697, 1329-1621, 1336-1546,
	1341-1559, 1342-1688, 1348-1684, 1348-1697, 1349-1602, 1350-1561, 1350-1686, 1352-1697, 1353-1686, 1354-1690, 1355-1630, 1355-1690, 1358-1574, 1358-1692, 1359-1648, 1372-1506, 1376-1614, 1381-1688, 1385-1697, 1386-1668, 1387-1688, 1388-1625, 1390-1688, 1390-1697, 1392-1598, 1392-1670, 1392-1673, 1392-1690, 1394-1688, 1395-1606, 1396-1686, 1399-1563, 1403-1669, 1406-1692, 1406-1697, 1407-1677, 1409-1697, 1412-1566, 1412-1697, 1418-1688, 1426-1688, 1427-1688, 1428-1689, 1430-1688, 1431-1641, 1431-1697, 1435-1697, 1443-1689, 1445-1697, 1446-1690, 1447-1688, 1459-1658, 1469-1685, 1512-1692, 1515-1688, 1515-1697, 1521-1697, 1522-1688, 1531-1697, 1533-1697, 1547-1687, 1550-1697, 1555-1697, 1565-1688, 1568-1643, 1579-1690, 1579-1697, 1582-1697, 1596-1697, 1616-1687, 1621-1690, 1632-1697
77/7509963CB1 1605	1-185, 1-265, 1-1603, 3-280, 3-379, 28-691, 28-796, 28-844, 28-849, 28-856, 29-839, 29-851, 29-856, 36-305, 44-229, 44-311, 45-292, 46-306, 46-538, 47-285, 49-194, 49-317, 50-299, 51-266, 52-679, 53-329, 54-344, 55-304, 58-284, 58-318, 58-326, 58-535, 58-664, 59-282, 59-473, 59-610, 60-623, 61-318, 61-405, 61-546, 61-571, 61-690, 62-334, 63-177, 63-305, 63-322, 64-285, 64-302, 64-321, 64-322, 64-356, 64-368, 64-408, 65-167, 65-214, 65-263, 65-301, 65-302, 65-316, 65-335, 65-337, 65-420, 65-641, 65-683, 65-715, 66-256, 66-309, 66-318, 66-337, 66-346, 66-388, 66-851, 67-317, 67-349, 68-351, 68-356, 70-506, 70-554, 70-564, 70-661, 70-739, 71-316, 71-772, 72-238, 72-554, 73-296, 73-308, 73-326, 73-335, 73-337, 73-341, 73-357, 73-551, 73-602, 73-664, 73-728, 74-204, 74-313, 74-316, 74-317, 74-319, 74-321, 74-324, 74-331, 74-338, 74-340, 74-341, 74-344, 74-347, 74-348, 74-351, 74-371, 74-375, 74-382, 74-724, 75-272, 75-326, 75-365, 75-728, 76-318, 76-405, 77-680, 77-698, 78-198, 80-324, 80-728, 80-762, 82-318, 951-1035, 951-1059, 954-1012,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	82-340, 82-677, 83-697, 84-306, 84-342, 84-423, 84-756, 88-399, 89-320, 90-723, 91-644, 93-623, 95-275, 95-295, 95-308, 95-345, 95-381, 95-550, 95-553, 95-666, 97-352, 114-397, 116-357, 117-499, 147-415, 162-435, 178-432, 178-441, 187-461, 188-452, 203-416, 220-636, 223-602, 238-592, 243-592, 243-606, 246-508, 247-507, 260-461, 272-542, 298-506, 299-477, 299-515, 322-856, 328-585, 341-577, 353-648, 356-634, 361-615, 387-656, 387-856, 390-632, 390-850, 406-669, 406-696, 409-744, 410-541, 414-662, 421-703, 426-638, 426-656, 440-757, 442-717, 445-703, 445-705, 451-644, 451-702, 453-607, 453-717, 459-693, 459-708, 461-685, 461-726, 466-608, 466-724, 470-775, 471-687, 472-888, 473-780, 485-778, 492-786, 492-791, 503-744, 503-791, 504-740, 509-760, 514-786, 515-765, 518-773, 526-797, 527-817, 557-824, 559-633, 561-812, 569-799, 596-856, 602-847, 622-815, 627-808, 636-909, 644-857, 655-818, 669-866, 673-805, 723-856, 753-1166, 753-1199, 879-1519, 881-1129, 951-991, 951-1028, 951-1029, 954-1022, 1212-1561, 1212-1562, 1537-1605
78/7505089CB1 1576	1-199, 1-230, 1-236, 1-286, 1-292, 1-321, 1-1448, 2-503, 3-248, 3-253, 3-709, 8-189, 8-234, 8-248, 8-703, 8-723, 12-280, 12-297, 19-270, 19-279, 20-233, 21-265, 22-154, 22-250, 22-271, 22-351, 23-291, 26-290, 29-281, 36-252, 37-315, 42-732, 45-252, 45-277, 45-286, 45-294, 45-297, 45-307, 46-303, 46-422, 47-310, 49-168, 49-332, 49-342, 51-199, 51-301, 52-284, 58-182, 58-197, 58-363, 60-397, 61-307, 61-308, 61-312, 61-320, 61-449, 62-275, 62-335, 62-352, 63-291, 64-411, 65-228, 65-314, 65-338, 66-185, 66-200, 66-254, 66-299, 66-342, 66-344, 66-347, 66-627, 66-676, 67-257, 67-291, 67-314, 67-331, 67-346, 67-347, 67-352, 67-362, 67-371, 67-376, 67-386, 68-330, 68-448, 68-747, 69-225, 69-312, 69-345, 69-350, 70-344, 70-351; 72-331, 72-359, 75-318, 75-499, 76-321, 76-344, 77-341, 79-166, 79-370, 80-346, 81-257, 82-323, 82-476, 83-319, 83-364, 83-386, 84-360, 84-362, 85-344, 88-325, 89-503, 92-318, 94-356, 94-376, 96-466, 97-331, 97-372, 98-370, 99-705, 104-353, 107-755, 111-281, 111-373, 112-385, 113-278, 113-360,
	115-358, 134-352, 167-363, 179-503, 198-396, 214-451, 214-462, 225-328, 225-446, 228-442, 235-503, 248-616, 254-456, 254-502, 265-503, 269-417, 275-503, 281-503, 284-478, 315-440, 319-588, 319-606, 339-486, 344-494, 482-958, 483-809, 493-696, 504-730, 504-733, 504-768, 504-956, 504-965, 504-975, 504-1078, 504-1097, 506-744, 508-656, 508-1143, 517-1088, 518-856, 519-854, 520-722, 520-834, 521-972, 526-802, 527-768, 531-753, 531-1036, 534-796, 535-774, 537-756, 537-789, 541-765, 549-838, 552-819, 557-772, 561-1156, 563-839, 564-841, 564-1134, 571-829, 571-830, 571-1148, 571-1305, 572-833, 573-1088, 573-1094, 580-817, 580-1000, 580-1111, 582-1220, 584-1171, 586-806, 594-796, 594-854, 594-867, 594-879, 594-883, 594-922, 600-972, 602-1203, 605-742, 608-1309, 611-871, 611-912, 611-1238, 612-714, 612-805, 613-1074, 614-913, 614-954, 620-983, 630-1085, 632-896, 632-927, 633-876, 637-1054, 637-1069, 637-1210, 638-1309, 640-1237, 645-913, 645-1306, 645-1334, 647-902, 652-1272, 652-1282, 658-972, 662-888, 662-929,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	664-1423, 669-1277, 670-931, 671-834, 673-841, 673-940, 677-1185, 678-1165, 682-947, 682-988, 683-1257, 685-1171, 686-1399, 687-957, 688-914, 692-936, 692-943, 693-1262, 698-1050, 699-1203, 701-983, 703-992, 706-970, 706-992, 706-1174, 708-956, 708-960, 709-1326, 711-1317, 714-913, 714-940, 714-1007, 714-1203, 720-1354, 721-1074, 721-1183, 727-1002, 727-1012, 727-1245, 729-1419, 730-1132, 730-1141, 730-1176, 731-993, 731-1330, 736-1325, 742-1287, 749-1128, 749-1281, 750-1048, 750-1111, 750-1323, 751-1362, 751-1390, 753-976, 753-1194, 753-1291, 753-1323, 754-1228, 757-1015, 771-942, 771-965, 771-970, 771-1028, 771-1032, 771-1055, 771-1324, 777-1174, 779-1038, 779-1041, 782-1400, 784-1408, 785-1070, 789-1037, 789-1054, 790-1111, 790-1356, 792-1007, 792-1255, 798-1388, 799-1458, 803-1051, 803-1067, 804-1086, 807-1061, 807-1331, 808-1308, 810-982, 810-1067, 810-1074, 810-1475, 812-1394, 813-1215, 813-1247, 813-1366, 815-1076, 815-1285, 819-992, 820-1448, 821-1297, 822-1060, 824-1454, 825-1364, 825-1366,
	827-1422, 829-1107, 833-1087, 834-1233, 835-961, 837-1329, 840-1306, 841-1165, 843-1461, 844-1324, 847-1143, 851-1037, 851-1385, 851-1421, 854-1088, 854-1100, 858-1481, 860-1134, 861-1215, 864-1128, 865-1149, 867-1124, 867-1460, 868-1139, 870-1382, 872-984, 875-1132, 878-1102, 879-1467, 880-1343, 881-1088, 882-1331, 887-1173, 893-1183, 894-1219, 894-1221, 894-1463, 895-972, 902-1153, 903-1179, 903-1507, 907-1102, 907-1126, 907-1451, 908-1160, 910-1163, 910-1366, 910-1387, 913-1100, 914-1455, 915-1161, 915-1171, 918-1174, 918-1191, 918-1366, 918-1469, 920-1227, 920-1232, 922-1168, 924-1322, 929-1177, 929-1452, 930-1182, 931-1244, 933-1242, 934-1206, 936-992, 936-1209, 936-1392, 936-1460, 939-1207, 940-1145, 941-1187, 944-1212, 944-1232, 949-1184, 951-1313, 955-1181, 956-1226, 957-1227, 959-1210, 962-1226, 964-1097, 964-1218, 964-1222, 966-1196, 967-1226, 978-1366, 986-1398, 988-1398, 990-1261, 996-1099, 1000-1325, 1002-1222, 1002-1275, 1002-1371, 1003-1272, 1009-1309, 1020-1311, 1022-1302, 1026-1298,
	1026-1516, 1027-1283, 1047-1278, 1061-1307, 1064-1319, 1069-1575, 1072-1307, 1083-1322, 1084-1313, 1084-1331, 1084-1332, 1084-1576, 1092-1296, 1093-1339, 1093-1446, 1104-1360, 1110-1378, 1111-1426, 1112-1370, 1113-1325, 1120-1384, 1127-1263, 1128-1354, 1128-1363, 1128-1414, 1147-1384, 1147-1391, 1183-1447, 1184-1417, 1184-1434, 1185-1412, 1185-1440, 1185-1450, 1185-1454, 1197-1443, 1198-1422, 1202-1436, 1209-1299, 1209-1441, 1209-1445, 1226-1370, 1231-1401, 1231-1475, 1237-1454, 1288-1445
79/7510139CB1 1695	1-253, 1-380, 1-441, 1-645, 1-679, 1-737, 1-775, 1-792, 1-830, 1-1685, 2-630, 2-719, 5-375, 5-432, 6-613, 71-325, 101-622, 147-399, 152-434, 155-397, 231-372, 231-436, 247-509, 256-614, 267-537, 277-502, 277-557, 291-818, 307-527, 318-564, 318-570, 319-557, 323-594, 325-567, 343-616, 358-596, 369-539, 375-517, 392-652, 409-691, 462-699, 467-746, 481-1111, 559-1062, 588-812, 588-830, 711-1138, 788-1039, 799-931, 799-1051, 877-1143, 938-1281, 940-1648, 972-1674, 1003-1292, 1003-1393, 1017-1632, 1038-1299, 1049-1352, 1049-1536, 1049-1537, 1049-1654, 1049-1658, 1055-1275, 1055-1293, 1066-1695, 1103-1402, 1120-1695, 1131-1307, 1147-1385, 1163-1398, 1203-1695, 1228-1640, 1237-1695, 1259-1644, 1279-1501, 1281-1641, 1289-1666, 1294-1695, 1309-1695, 1317-1695, 1318-1671, 1336-1650, 1340-1662, 1345-1684, 1345-1689, 1355-1439, 1365-1679, 1369-1603, 1377-1645, 1377-1650, 1377-1657, 1381-1695, 1383-1494, 1386-1690, 1413-1695, 1420-1695, 1422-1582, 1442-1662, 1503-1695, 1520-1649, 1526-1690, 1530-1692, 1584-1695

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
80/7505053CB1 1370	1-101, 1-112, 1-124, 1-135, 1-139, 1-148, 1-149, 1-154, 1-159, 1-1358, 12-159, 25-77, 43-159, 49-159, 57-144, 60-140, 61-159, 62-159, 67-159, 112-324, 112-342, 157-351, 157-367, 157-388, 157-389, 157-423, 157-566, 157-735, 157-803, 157-874, 158-453, 163-461, 167-336, 169-591, 171-341, 186-455, 187-979, 188-484, 193-495, 205-318, 211-430, 212-282, 212-430, 213-883, 222-714, 230-883, 234-539, 242-915, 244-915, 254-710, 256-915, 262-623, 269-915, 271-751, 277-811, 287-915, 291-811, 316-561, 317-501, 333-627, 333-839, 333-887, 333-892, 340-915, 358-632, 359-599, 359-915, 360-610, 363-625, 373-632, 374-934, 377-643, 391-839, 393-915, 395-915, 413-683, 415-611, 419-624, 420-611, 422-704, 430-590, 444-700, 444-741, 445-843, 446-677, 449-721, 456-725, 457-744, 457-750, 464-901, 467-740, 468-623, 473-727, 473-739, 473-862, 473-1022, 475-768, 480-883, 486-901, 487-654, 488-949, 493-927, 497-1047, 498-889, 498-1156, 499-759, 503-641, 503-1119, 504-788, 506-1357, 511-896, 511-907, 518-889, 532-1338, 534-1203, 537-834,
	539-1096, 540-803, 553-672, 564-1118, 570-663, 571-816, 576-794, 576-872, 590-1123, 598-1251, 604-859, 611-1228, 612-816, 622-863, 631-1308, 632-937, 633-953, 641-912, 645-1307, 656-908, 656-1055, 656-1062, 657-859, 668-976, 680-1319, 680-1347, 684-1346, 686-949, 691-1279, 698-983, 699-996, 699-1011, 699-1035, 700-903, 714-987, 732-958, 757-1016, 764-1023, 765-1045, 772-1029, 773-1042, 773-1085, 782-1051, 786-1047, 795-1115, 795-1358, 798-1114, 803-1108, 808-1091, 810-1060, 813-1347, 814-1337, 815-1076, 816-1065, 816-1358, 817-1153, 818-1040, 818-1047, 829-1106, 833-1068, 852-1158, 862-1155, 865-1089, 865-1119, 870-1145, 874-1358, 875-1045, 876-1156, 878-1045, 885-1052, 887-1045, 891-1357, 897-1370, 898-1045, 899-1047, 911-1358, 915-1073, 916-1358, 924-1162, 924-1365, 931-1047, 932-1047, 935-1047, 935-1185, 943-1047, 945-1133, 945-1296, 945-1367, 946-1045, 950-1244, 951-1047, 952-1047, 952-1175, 958-1304, 959-1045, 960-1047, 970-1047, 971-1047, 971-1189, 979-1102, 980-1045, 984-1231, 993-1045, 994-1045,
	997-1176, 1019-1076, 1084-1118, 1218-1252, 1227-1291
81/7511116CB1 2844	1-324, 1-460, 1-461, 1-488, 1-526, 1-532, 1-591, 1-752, 1-2570, 235-700, 323-873, 382-1018, 437-1080, 467-1044, 582-1185, 690-1151, 727-1219, 750-1428, 805-1238, 979-1615, 992-1578, 1116-1589, 1120-1736, 1137-1712, 1152-1776, 1174-1789, 1199-1763, 1301-1780, 1305-1787, 1386-1959, 1397-2011, 1411-2045, 1509-2056, 1519-1726, 1527-1701, 1551-2090, 1561-1725, 1662-2158, 1665-2106, 1680-2170, 1704-2257, 1732-2221, 1749-2230, 1769-2319, 1792-2286, 1808-2471, 1818-2344, 1865-2248, 1924-2574, 1957-2555, 1997-2209, 2033-2597, 2057-2619, 2087-2582, 2093-2565, 2180-2844, 2233-2566
82/7511175CB1 1174	1-172, 1-312, 1-383, 1-490, 1-501, 1-503, 1-562, 1-1162, 97-563, 314-851, 318-851, 504-1174, 563-1110, 587-1174, 689-1174

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
83/7504660CB1 1153	1-228, 1-1153, 17-228, 24-119, 26-201, 80-275, 81-261, 81-316, 81-356, 81-520, 81-583, 81-704, 81-713, 228-472, 228-668, 228-735, 229-609, 230-668, 231-730, 240-511, 243-463, 245-509, 253-544, 254-508, 258-449, 258-466, 265-512, 267-427, 268-520, 284-707, 286-524, 287-510, 300-522, 300-542, 307-527, 308-592, 318-567, 324-558, 325-591, 326-589, 328-470, 328-713, 330-575, 331-587, 331-597, 331-652, 331-653, 339-506, 341-533, 346-693, 350-558, 351-482, 362-716, 367-713, 378-585, 379-598, 379-607, 379-608, 379-616, 379-621, 379-635, 379-658, 380-633, 386-716, 389-612, 390-669, 393-619, 393-652, 393-654, 394-598, 396-626, 398-615, 399-648, 400-539, 401-660, 403-696, 415-654, 415-655, 416-673, 420-643, 420-664, 420-665, 422-668, 423-664, 436-622, 438-705, 438-717, 439-696, 440-668, 445-707, 452-660, 452-690, 453-675, 462-589, 475-711, 476-722, 512-724
84/7504681CB1 3628	1-253, 1-3345, 17-228, 21-315, 49-768, 245-893, 347-939, 383-632, 384-681, 473-731, 473-759, 473-899, 473-923, 473-942, 473-987, 473-1009, 473-1044, 473-1070, 473-1082, 473-1094, 473-1104, 562-1190, 619-822, 671-1166, 672-1244, 684-1256, 721-984, 726-1293, 731-1566, 741-1237, 774-1362, 830-1405, 872-1362, 878-1468, 879-1096, 896-1261, 984-1293, 989-1652, 995-1449, 995-1488, 998-1437, 998-1611, 1009-1710, 1023-1691, 1084-1360, 1108-1386, 1165-1420, 1165-1606, 1174-1717, 1178-1668, 1229-1406, 1238-1485, 1240-1496, 1282-1737, 1286-1737, 1300-1639, 1307-1944, 1336-2050, 1338-1747, 1369-1640, 1379-1751, 1391-1662, 1425-1683, 1425-1715, 1455-1896, 1461-1720, 1484-1997, 1490-2056, 1506-2013, 1563-1818, 1582-1729, 1594-1867, 1606-2262, 1639-2110, 1651-1921, 1695-1982, 1695-2227, 1749-1944, 1749-2050, 1778-1991, 1778-2000, 1778-2018, 1779-2325, 1785-2013, 1785-2078, 1852-2136, 1877-2154, 1902-2174, 1906-2162, 2010-2171, 2042-2724, 2043-2629, 2045-2319, 2065-2317, 2073-2351, 2088-2384, 2105-2367, 2105-2668, 2142-2964,
	2151-2423, 2151-2601, 2151-2717, 2194-2789, 2201-2450, 2223-3010, 2227-2866, 2234-2496, 2235-2462, 2235-2715, 2253-2542, 2253-2812, 2256-2909, 2285-2910, 2309-2636, 2322-2665, 2340-2509, 2340-2545, 2366-3008, 2367-2615, 2368-2625, 2384-2659, 2384-2982, 2401-3338, 2413-3338, 2425-2687, 2429-2757, 2433-2999, 2437-2615, 2437-2622, 2437-2771, 2445-3336, 2451-2705, 2451-2720, 2487-2728, 2487-3107, 2499-2638, 2500-2753, 2501-2844, 2501-2877, 2501-2883, 2501-2984, 2504-2984, 2509-2760, 2510-3338, 2513-3338, 2515-2817, 2515-3033, 2523-3338, 2535-3336, 2539-2769, 2539-2906, 2546-3010, 2552-3182, 2553-2802, 2553-2825, 2553-3085, 2569-3408, 2573-3022, 2579-2876, 2590-2786, 2591-3338, 2592-3338, 2593-3338, 2594-3337, 2598-2877, 2599-3291, 2600-3325, 2614-2963, 2616-2771, 2616-2803, 2622-2872, 2626-2906, 2627-3122, 2629-3338, 2633-3445, 2634-2790, 2640-3078, 2654-3399, 2660-2901, 2663-3126, 2671-3225, 2675-2894, 2679-3009, 2679-3126, 2681-3338, 2683-3338, 2690-2916, 2691-3041, 2716-2972, 2716-3338, 2730-3003, 2738-2970, 2745-2979, 2745-3343, 2752-3323, 2763-3338, 2770-3227,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	2770-3236, 2771-3199, 2772-2832, 2772-3009, 2772-3122, 2776-3338, 2779-3304, 2780-3045, 2782-2940, 2785-3295, 2786-3389, 2792-3337, 2794-3295, 2797-3058, 2804-3189, 2811-3301, 2811-3356, 2817-3539, 2818-3440, 2826-3308, 2833-3122, 2833-3356, 2844-3102, 2851-3226, 2858-3509, 2862-3464, 2862-3628, 2864-3147, 2866-3464, 2867-3628, 2871-3263, 2876-3154, 2878-3472, 2879-3474, 2881-3464, 2893-3181, 2908-3345, 2910-3387, 2910-3473, 2916-3167, 2916-3608, 2921-3177, 2943-3409, 2944-3475, 2951-3330, 2955-3429, 2959-3476, 2988-3220, 2993-3473, 2994-3252, 2996-3330, 2998-3476, 3000-3268, 3003-3330, 3006-3480, 3011-3500, 3018-3263, 3023-3235, 3025-3189, 3027-3464, 3029-3330, 3030-3330, 3043-3330, 3046-3445, 3063-3299, 3063-3330, 3063-3341, 3075-3498, 3080-3476, 3081-3330, 3082-3330, 3082-3343, 3093-3577, 3096-3464, 3099-3330, 3105-3628, 3116-3395, 3117-3330, 3120-3430, 3123-3524, 3127-3480, 3135-3480, 3135-3567, 3148-3323, 3158-3556, 3169-3628, 3172-3628, 3174-3422, 3174-3432, 3220-3459, 3224-3330, 3230-3485, 3239-3330, 3243-3628, 3247-3341, 3260-3514, 3260-3628, 3262-3330, 3343-3394, 3356-3398, 3357-3398, 3513-3538
85/7506472CB1 1160	1-1144, 13-599, 13-633, 108-472, 122-527, 140-498, 140-628, 140-666, 140-770, 140-782, 140-813, 143-824, 146-679, 146-742, 149-743, 150-750, 152-666, 157-666, 185-666, 217-808, 219-1004, 227-641, 277-848, 281-1094, 286-557, 337-633, 339-922, 339-993, 349-999, 368-996, 376-814, 378-999, 382-999, 385-1138, 386-990, 391-1092, 406-579, 408-925, 472-990, 472-1022, 486-990, 523-1120, 524-1034, 524-1120, 526-1092, 532-1128, 551-759, 551-1079, 568-1085, 592-1160, 772-1160, 802-836, 874-1105, 874-1128, 908-993
86/7506483CB1 1410	1-199, 1-230, 1-236, 1-265, 1-285, 1-1282, 3-248, 3-253, 3-284, 8-234, 8-248, 12-277, 12-280, 12-285, 13-284, 15-284, 19-270, 19-279, 19-284, 20-233, 20-284, 21-265, 22-154, 22-250, 22-271, 23-285, 26-284, 29-281, 36-252, 37-284, 45-252, 45-277, 45-284, 45-285, 46-285, 49-168, 51-199, 51-284, 52-284, 58-182, 58-197, 61-281, 61-284, 62-261, 62-275, 63-285, 65-228, 65-285, 66-185, 66-200, 66-254, 66-276, 66-284, 66-285, 66-287, 67-257, 67-284, 67-285, 68-285, 69-225, 70-278, 70-284, 72-284, 75-285, 76-885, 76-981, 77-731, 79-166, 81-257, 82-285, 88-425, 88-485, 88-623, 88-646, 88-647, 88-701, 88-710, 90-643, 111-281, 113-278, 149-313, 165-659, 203-706, 284-368, 284-447, 284-449, 284-480, 284-512, 284-520, 284-521, 284-551, 284-560, 284-657, 284-696, 284-761, 284-894, 285-602, 289-528, 290-947, 291-579, 293-602, 298-1001, 302-521, 309-931, 311-564, 311-567, 311-799, 312-627, 313-564, 313-578, 314-790, 316-549, 325-809, 328-912, 337-506, 338-530, 338-643, 338-792, 342-490, 342-977, 346-422, 351-922, 352-690,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
86 - Cont'd	353-688, 354-556, 354-668, 355-806, 360-636, 361-602, 365-587, 365-870, 368-630, 369-608, 371-590, 371-623, 375-599, 383-672, 386-653, 391-606, 395-990, 397-673, 398-675, 398-968, 405-663, 405-664, 405-982, 405-1139, 406-667, 407-922, 407-928, 414-651, 414-834, 414-945, 416-1054, 418-1005, 420-640, 428-630, 428-688, 428-701, 428-717, 428-756, 434-806, 436-1037, 439-576, 442-1143, 445-705, 445-746, 445-1072, 446-548, 446-639, 447-908, 448-747, 448-788, 454-817, 464-919, 466-730, 466-761, 467-710, 471-888, 471-903, 471-1044, 472-1143, 474-1071, 479-747, 479-1140, 479-1168, 481-736, 486-1106, 486-1116, 492-806, 496-722, 496-763, 498-1257, 503-1111, 504-765, 505-668, 507-675, 507-774, 511-1019, 512-999, 516-781, 516-822, 517-1091, 519-1005, 520-1233, 521-791, 522-748, 526-770, 526-777, 527-1096, 532-884, 533-1037, 535-817, 537-826, 540-804, 540-826, 540-1008, 542-790, 542-794, 543-1160, 545-1151, 548-747, 548-774, 548-841, 548-1037, 554-1188, 555-908, 555-1017, 561-836, 561-846, 561-1079, 563-1253, 564-966,
	564-975, 565-827, 565-1164, 570-1159, 576-1121, 583-962, 583-1115, 584-945, 584-1157, 585-857, 585-1196, 585-1224, 587-810, 587-1028, 587-1125, 587-1157, 588-1062, 591-849, 605-776, 605-799, 605-804, 605-866, 605-889, 605-1158, 611-1008, 613-872, 613-875, 616-1234, 618-1242, 619-904, 623-871, 623-888, 624-945, 624-1190, 626-841, 626-1089, 632-1222, 633-1292, 637-885, 637-901, 641-895, 641-1165, 642-1142, 644-816, 644-901, 644-908, 644-1309, 646-1228, 647-1049, 647-1081, 647-1200, 649-910, 649-1119, 653-826, 654-1282, 655-1131, 658-1288, 659-1198, 659-1200, 661-1256, 663-941, 667-921, 668-1067, 669-795, 671-1163, 674-1140, 675-999, 677-1295, 678-1158, 681-977, 685-871, 685-1219, 685-1255, 688-922, 688-934, 692-1315, 694-968, 695-1049, 698-962, 699-983, 701-958, 701-1294, 702-973, 704-1216, 706-818, 709-966, 712-936, 713-1301, 714-1177, 715-922, 716-1165, 721-1007, 723-966, 727-1017, 728-1053, 728-1055, 728-1297, 729-806, 736-987, 737-891, 737-1013, 737-1341, 741-936, 741-960, 741-1285, 742-994, 744-997,
	744-1200, 744-1221, 747-934, 748-1289, 749-995, 752-1008, 752-1025, 752-1200, 752-1303, 754-1061, 754-1066, 756-1002, 758-1156, 763-1011, 763-1286, 764-1016, 765-1078, 767-1076, 768-1040, 770-826, 770-1043, 770-1226, 770-1294, 773-1041, 774-979, 775-1021, 778-1046, 778-1066, 783-1018, 785-1147, 789-1015, 790-1060, 791-1061, 793-1044, 796-1060, 798-931, 798-1052, 798-1056, 800-1030, 801-1060, 812-1200, 820-1232, 822-1232, 824-1095, 830-933, 834-1159, 836-1056, 836-1109, 836-1205, 837-1106, 843-1143, 854-1145, 856-1136, 860-1132, 860-1350, 861-1117, 881-1112, 895-1141, 898-1153, 903-1409, 906-1141, 917-1156, 918-1147, 918-1165, 918-1166, 918-1410, 926-1130, 927-1173, 927-1280, 938-1194, 944-1212, 945-1260, 946-1204, 947-1159, 954-1218, 961-1097, 962-1188, 962-1197, 962-1248, 981-1218, 981-1225, 1017-1281, 1018-1251, 1019-1246, 1019-1274, 1019-1284, 1019-1288, 1031-1277, 1032-1256, 1036-1270, 1043-1133, 1043-1279, 1065-1235, 1065-1309, 1071-1288, 1090-1243

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
87/7506525CB1 2481	1-475, 1-2481, 3-673, 27-288, 83-252, 83-354, 83-535, 108-218, 108-300, 108-339, 108-462, 108-546, 108-573, 108-582, 109-379, 109-452, 109-541, 110-182, 112-182, 112-220, 112-223, 115-223, 115-530, 120-339, 120-613, 124-339, 131-339, 238-685, 238-689, 339-544, 385-572, 385-586, 386-671, 549-786, 549-790, 549-807, 645-936, 645-1173, 698-1332, 709-966, 875-1346, 964-1346, 982-1342, 992-1342, 1004-1340, 1039-1297, 1151-1346, 1192-1828, 1242-1828, 1336-1582, 1344-1619, 1344-1664, 1364-1750, 1378-1750, 1396-1957, 1413-2061, 1448-2038, 1480-1750, 1529-1750, 1547-1659, 1585-2141, 1593-1748, 1600-2271, 1620-1889, 1797-1928, 1798-2082, 1800-2360, 1857-2481
88/7506549CB1 3808	1-575, 1-647, 1-704, 1-726, 1-739, 1-769, 1-807, 1-3435, 2-516, 2-669, 25-666, 321-898, 632-959, 1332-1533, 1386-2124, 1395-2124, 1402-1671, 1552-2124, 1590-1853, 1608-2118, 1623-2112, 1623-2124, 1702-1996, 1702-2206, 1707-2198, 1842-2081, 1942-2331, 2184-2770, 2211-2422, 2211-2483, 2211-2651, 2211-2663, 2211-2677, 2211-2701, 2211-2715, 2211-2775, 2211-2796, 2211-2830, 2211-2862, 2211-2877, 2213-2909, 2222-2822, 2326-2897, 2424-2821, 2452-2796, 2599-3409, 2631-2917, 2712-2910, 2805-3408, 2927-3443, 2955-3406, 2957-3409, 2958-3443, 3089-3808, 3103-3422, 3117-3420, 3242-3420
89/7506683CB1 5238	1-641, 1-649, 8-152, 8-555, 8-594, 8-648, 8-649, 8-5238, 13-537, 268-1103, 273-1103, 310-1103, 322-1103, 1037-1500, 1049-1506, 1057-1513, 1070-1513, 1083-1489, 1103-1513, 1158-1513, 1239-1500
90/7510814CB1 3200	1-851, 1-3200, 109-398, 174-840, 298-737, 348-819, 350-1101, 357-1080, 364-1017, 370-1019, 389-1172, 413-1032, 417-1083, 420-996, 433-1147, 448-1050, 455-769, 476-769, 477-765, 477-769, 492-1403, 512-1363, 533-1171, 541-769, 558-1181, 559-1353, 591-1153, 614-1279, 629-1472, 647-1543, 662-1335, 663-1515, 682-1067, 682-1169, 682-1315, 682-1327, 682-1329, 682-1352, 682-1375, 682-1384, 682-1399, 682-1404, 682-1424, 682-1430, 682-1434, 682-1443, 682-1470, 682-1471, 682-1472, 682-1478, 682-1484, 682-1488, 682-1500, 682-1503, 682-1516, 682-1523, 682-1529, 682-1530, 682-1533, 682-1534, 682-1546, 682-1549, 682-1550, 682-1553, 682-1557, 682-1558, 682-1561, 682-1563, 682-1568, 682-1572, 682-1574, 682-1578, 682-1582, 682-1583, 682-1587, 682-1588, 682-1597, 682-1599, 682-1600, 682-1602, 682-1614, 682-1621, 682-1643, 682-1661, 682-1691, 682-1693, 684-1552, 684-1612, 685-887, 686-1243, 694-769, 747-1083, 751-1516, 778-1497, 781-1507, 787-1555, 789-1427, 814-1528, 841-1441, 863-1643, 866-1571, 871-1372, 914-1050, 932-1176,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
90 - Cont'd	936-1094, 969-1735, 985-1811, 1016-1565, 1018-1127, 1037-1314, 1083-1767, 1091-1728, 1110-1707, 1130-1929, 1132-1916, 1137-1370, 1137-1384, 1137-1509, 1137-1593, 1137-1607, 1137-1666, 1137-1689, 1137-1711, 1137-1714, 1137-1739, 1137-1818, 1140-1703, 1164-1715, 1171-1867, 1188-1344, 1188-1840, 1190-1846, 1210-1612, 1216-2215, 1217-1818, 1221-2216, 1225-1930, 1228-1827, 1232-2216, 1234-2215, 1234-2216, 1239-2216, 1240-2216, 1248-1681, 1248-2215, 1250-2216, 1251-1821, 1253-2216, 1254-1416, 1256-1735, 1257-2216, 1258-2216, 1261-2216, 1265-2216, 1274-2216, 1278-2216, 1280-2216, 1282-1404, 1282-1947, 1284-2216, 1285-2216, 1286-2215, 1286-2216, 1288-2214, 1290-2212, 1290-2216, 1294-2030, 1294-2216, 1297-2216, 1299-2216, 1302-2216, 1303-1925, 1306-2216, 1307-2216, 1309-1632, 1317-2216, 1318-1985, 1319-1930, 1321-1919, 1323-2215, 1323-2216, 1329-2216, 1331-2216, 1333-2216, 1335-2216, 1339-2216, 1342-2216, 1343-2019, 1351-2216, 1352-2216, 1353-1853, 1353-2214, 1353-2216, 1355-2004, 1357-2214, 1357-2216, 1360-2216,
	1362-2216, 1363-1841, 1364-2216, 1371-2025, 1374-1958, 1374-2216, 1378-2017, 1378-2216, 1379-2216, 1380-2216, 1383-2216, 1385-2216, 1386-2216, 1389-2216, 1390-2216, 1391-2039, 1396-1979, 1402-2091, 1408-2216, 1409-1974, 1409-2215, 1409-2216, 1410-2214, 1410-2216, 1411-2216, 1412-2047, 1417-2216, 1418-2216, 1419-2216, 1421-2216, 1422-2216, 1423-2216, 1430-2216, 1431-2216, 1432-2216, 1434-2216, 1435-2216, 1436-2029, 1439-1684, 1439-1761, 1439-2060, 1440-1876, 1441-1705, 1441-1713, 1441-1728, 1441-1902, 1441-1906, 1441-1912, 1441-1974, 1441-2028, 1441-2066, 1441-2215, 1442-2190, 1444-2216, 1445-2080, 1445-2216, 1448-1938, 1448-2212, 1448-2216, 1452-2216, 1453-2212, 1453-2216, 1455-2216, 1456-1987, 1458-2216, 1459-2216, 1460-2033, 1461-2216, 1466-2216, 1471-2069, 1471-2216, 1478-2216, 1479-2175, 1479-2216, 1481-2216, 1484-2216, 1485-2216, 1487-2216, 1490-2216, 1493-2216, 1494-2047, 1498-2174, 1501-2216, 1511-2216, 1514-2212, 1516-2166, 1519-2203, 1524-2216, 1525-2216, 1527-2216, 1535-2216, 1540-2250, 1547-2216,
	1549-2142, 1550-2216, 1551-1925, 1552-2107, 1552-2216, 1553-2119, 1561-2035, 1568-2138, 1573-2167, 1574-2110, 1575-2216, 1577-2206, 1578-2216, 1584-2310, 1586-2216, 1587-2211, 1589-2103, 1596-2125, 1600-2123, 1600-2261, 1610-1954, 1612-2165, 1634-1996, 1636-2213, 1640-2216, 1642-2216, 1644-2216, 1650-2216, 1664-2132, 1666-2299, 1670-2189, 1673-2284, 1673-2320, 1676-1916, 1677-2318, 1679-2216, 1686-1878, 1687-2344, 1690-1832, 1691-2173, 1695-2230, 1701-1843, 1701-2061, 1701-2206, 1701-2432, 1711-1990, 1713-2218, 1715-2035, 1715-2216, 1715-2299, 1717-2298, 1718-2184, 1726-2378, 1727-2356, 1728-2197, 1730-2028, 1730-2241, 1738-1965, 1738-2077, 1741-2203, 1745-2236, 1745-2327, 1755-2369, 1757-1865, 1762-2383, 1763-2292, 1774-2364, 1776-1902, 1781-2391, 1781-2432, 1789-2251, 1790-2399, 1793-2043, 1793-2050, 1793-2366, 1793-2590, 1797-2358, 1804-2212, 1808-1967, 1810-2459, 1811-2298, 1816-2224, 1817-2349, 1820-2354, 1820-2595, 1822-2338, 1822-2637, 1824-2418, 1825-2359, 1834-2419, 1834-2511, 1836-2466, 1841-2396,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
90 - Cont'd	1844-2383, 1849-2389, 1849-2418, 1852-2500, 1857-2119, 1861-2304, 1865-2487, 1869-2486, 1876-2385, 1877-2431, 1879-2512, 1881-2138, 1883-2506, 1891-2454, 1891-2538, 1892-2425, 1893-2601, 1895-2449, 1896-2399, 1896-2425, 1896-2652, 1899-2560, 1903-2480, 1912-2538, 1914-2084, 1917-2451, 1920-2533, 1920-2556, 1921-2500, 1924-2076, 1924-2519, 1926-2506, 1926-2542, 1934-2572, 1935-2719, 1943-2477, 1944-2473, 1944-2493, 1944-2497, 1946-2439, 1946-2575, 1947-2493, 1949-2519, 1949-2562, 1950-2489, 1951-2428, 1951-2481, 1952-2602, 1957-2276, 1959-2556, 1964-2527, 1966-2506, 1968-2589, 1974-2533, 1978-2216, 1980-2348, 1980-2533, 1980-2538, 1982-2506, 1983-2321, 1983-2632, 1990-2533, 1997-2547, 1998-2526, 2004-2553, 2007-2551, 2009-2590, 2015-2650, 2022-2125, 2024-2674, 2027-2744, 2030-2715, 2032-2665, 2033-2632, 2036-2631, 2036-2716, 2037-2608, 2038-2578, 2038-2646, 2042-2260, 2042-2375, 2042-2554, 2048-2674, 2053-2656, 2055-2577, 2055-2663, 2058-2234, 2058-2682, 2059-2282, 2062-2658, 2065-2761, 2074-2676, 2077-2658,
	2091-2634, 2092-2360, 2093-2422, 2093-2627, 2101-2644, 2107-2674, 2109-2652, 2117-2663, 2121-2721, 2124-2500, 2124-2735, 2139-2734, 2161-2791, 2162-2284, 2162-2730, 2167-2548, 2168-2717, 2168-2731, 2177-2714, 2178-2799, 2180-2767, 2185-2398, 2190-2482, 2190-2641, 2190-2696, 2191-2546, 2191-2903, 2195-2326, 2196-2721, 2197-2696, 2198-2717, 2200-2723, 2201-2696, 2202-2797, 2205-2759, 2207-2832, 2213-2616, 2217-2428, 2217-2673, 2217-2675, 2218-2765, 2219-2721, 2220-2733, 2220-2845, 2223-2709, 2224-2799, 2230-2726, 2232-2473, 2240-2716, 2240-2815, 2243-2811, 2246-2379, 2250-2601, 2251-2629, 2252-2631, 2259-2842, 2273-2819, 2275-2736, 2281-2893, 2283-2559, 2284-2416, 2285-2840, 2288-2652, 2294-2855, 2296-2740, 2305-2591, 2305-2872, 2307-2742, 2309-2812, 2314-2558, 2315-2908, 2318-2401, 2318-2608, 2318-2796, 2318-2826, 2318-2905, 2318-2907, 2318-2912, 2318-2927, 2318-2956, 2318-3001, 2324-2953, 2334-3126, 2335-3118, 2339-2682, 2358-2873, 2360-3049, 2364-2968, 2369-2593, 2372-2753, 2373-2874, 2381-2593, 2393-2762,
	2413-2908, 2427-2900, 2436-2652, 2439-3096, 2439-3200, 2446-2695, 2446-2925, 2450-2740, 2457-2761, 2461-3121, 2465-2732, 2465-3035, 2471-2988, 2473-2981, 2484-2751, 2493-2745, 2493-2794, 2493-3028, 2496-3019, 2508-3192, 2508-3200, 2516-2764, 2516-3177, 2517-2617, 2518-3033, 2525-2795, 2533-3200, 2547-3173, 2555-3200, 2557-2660, 2560-3200, 2561-2813, 2568-3038, 2570-3200, 2582-2728, 2613-2970, 2617-3183, 2623-2912, 2630-3196, 2631-3200, 2632-3001, 2639-2908, 2657-2912, 2657-2964, 2672-3163, 2686-2992, 2701-3199, 2702-3200, 2712-3199, 2726-2983, 2726-3200, 2729-2908, 2730-3200, 2731-3197, 2736-3002, 2738-3036, 2741-3002, 2749-3054, 2753-2924, 2766-3196, 2766-3198, 2767-3200, 2768-3200, 2774-3058, 2776-2983, 2780-3200, 2781-2983, 2786-3200, 2787-3197, 2792-3200, 2813-2948, 2816-3200, 2817-2983, 2822-2980, 2823-3200, 2825-3200, 2840-2974, 2864-3002, 2867-3036, 2873-3198, 2922-2983, 2922-3198, 2947-3078, 2950-3077, 2962-3182

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
91/7504727CB1 1839	1-288, 3-285, 3-300, 3-1829, 12-296, 20-506, 26-487, 65-610, 81-719, 119-742, 234-738, 239-538, 258-738, 292-626, 292-733, 292-742, 389-615, 423-675, 469-740, 469-742, 584-850, 605-962, 672-859, 752-1022, 752-1603, 753-1137, 780-1164, 781-1163, 781-1363, 782-1015, 800-1381, 855-1318, 898-1170, 899-1341, 904-1586, 904-1590, 904-1603, 904-1650, 904-1687, 910-999, 938-1079, 948-1175, 980-1541, 1006-1757, 1008-1206, 1008-1565, 1017-1298, 1038-1433, 1041-1431, 1056-1431, 1072-1263, 1077-1327, 1082-1390, 1092-1598, 1106-1831, 1122-1220, 1135-1826, 1139-1373, 1154-1833, 1165-1609, 1173-1811, 1173-1835, 1181-1815, 1200-1233, 1204-1816, 1207-1832, 1210-1724, 1215-1508, 1218-1460, 1219-1814, 1224-1403, 1227-1832, 1231-1393, 1244-1559, 1244-1749, 1261-1726, 1262-1581, 1280-1787, 1301-1559, 1305-1568, 1317-1789, 1319-1816, 1325-1604, 1359-1828, 1361-1833, 1363-1839, 1364-1800, 1369-1759, 1373-1833, 1375-1668, 1376-1818, 1376-1839, 1379-1839, 1381-1839, 1382-1835, 1383-1839, 1390-1839, 1393-1822, 1393-1839, 1397-1839,
	1405-1839, 1410-1829, 1415-1811, 1417-1672, 1419-1839, 1420-1472, 1436-1827, 1441-1827, 1448-1709, 1450-1828, 1451-1828, 1456-1823, 1458-1839, 1470-1839, 1478-1825, 1486-1827, 1496-1829, 1512-1829, 1513-1826, 1543-1823, 1564-1839, 1567-1839, 1592-1795, 1592-1834, 1658-1814, 1669-1827, 1686-1839, 1706-1822, 1723-1839
92/7506958CB1 2075	1-2075, 114-958, 481-944, 659-1174, 676-1595, 688-1595, 696-1595, 741-1591, 746-1591, 751-1595, 756-1595, 769-1595, 771-1595, 780-1595, 784-1595, 797-1278, 798-1595, 810-1595, 812-1594, 812-1595, 825-1595, 828-1595, 831-1595, 852-1591, 870-1595, 872-1591, 873-1594, 885-1595, 887-1595, 892-1591, 909-1296, 914-1296, 919-1591, 920-1591, 923-1595, 925-1591, 929-1595, 970-1535, 973-1595, 1004-1595, 1076-1591, 1077-1591, 1100-1595, 1354-1843, 1396-1806, 1396-1889, 1398-1653, 1611-1961
93/7505332CB1 1814	1-1701, 113-752, 208-518, 208-627, 208-759, 235-321, 235-473, 259-770, 262-539, 264-567, 265-854, 275-753, 282-541, 294-773, 303-501, 305-787, 316-503, 323-555, 323-584, 330-567, 330-1069, 347-821, 369-866, 371-900, 373-627, 373-632, 373-763, 402-783, 402-896, 408-931, 421-1049, 427-729, 427-971, 455-955, 456-783, 457-755, 458-781, 465-964, 476-1147, 496-834, 503-980, 515-1100, 533-840, 533-1066, 547-846, 558-1147, 562-1089, 576-773, 580-1134, 585-1140, 586-865, 586-1010, 587-1007, 595-993, 595-1224, 599-1121, 601-848, 602-861, 610-849, 611-960, 612-1142, 615-1232, 629-879, 631-951, 642-911, 645-1142, 654-1112, 655-1112, 665-1223, 671-941, 677-902, 684-922, 689-1298, 694-962, 702-980, 704-1204, 706-906, 708-876, 712-1023, 712-1215, 714-947, 722-1136, 722-1170, 722-1322, 741-867, 747-914, 750-1203, 758-1138, 758-1286, 759-992, 759-1189, 759-1222, 760-1013, 760-1407, 761-1336, 765-960, 765-1343, 767-1544, 770-1099, 782-1243, 787-1236, 791-893, 794-1236, 803-1095, 804-1238, 806-1236, 813-1083, 814-1112,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	818-1078, 818-1079, 818-1236, 822-1129, 828-1145, 832-979, 833-1131, 836-1091, 847-1370, 855-1130, 856-1537, 860-1556, 875-1107, 875-1462, 876-1133, 877-1236, 884-1230, 900-1544, 903-1336, 912-1490, 912-1524, 925-1207, 926-1192, 927-1230, 927-1235, 928-1470, 947-1533, 952-1199, 954-1243, 955-1470, 955-1514, 977-1464, 982-1429, 984-1229, 988-1470, 989-1470, 994-1465, 994-1539, 996-1236, 1001-1143, 1009-1274, 1013-1275, 1016-1331, 1017-1464, 1021-1570, 1022-1467, 1030-1470, 1044-1591, 1047-1464, 1048-1464, 1054-1469, 1055-1552, 1057-1467, 1060-1232, 1060-1252, 1062-1463, 1068-1470, 1076-1470, 1077-1236, 1080-1362, 1081-1600, 1083-1469, 1083-1582, 1088-1453, 1094-1469, 1099-1469, 1100-1375, 1100-1566, 1100-1591, 1105-1236, 1110-1554, 1111-1465, 1113-1588, 1113-1600, 1114-1597, 1115-1343, 1119-1597, 1121-1814, 1126-1600, 1127-1347, 1128-1303, 1128-1382, 1129-1453, 1131-1591, 1132-1814, 1144-1404, 1148-1469, 1148-1470, 1148-1584, 1148-1587, 1148-1590, 1150-1591, 1152-1600, 1153-1599, 1154-1592, 1155-1600,
	1157-1571, 1157-1589, 1159-1236, 1161-1592, 1163-1452, 1164-1464, 1165-1464, 1168-1592, 1168-1596, 1172-1445, 1173-1470, 1173-1814, 1176-1600, 1179-1591, 1180-1588, 1180-1758, 1182-1592, 1183-1592, 1184-1590, 1186-1367, 1186-1464, 1186-1584, 1195-1600, 1195-1814, 1198-1308, 1199-1814, 1201-1598, 1203-1598, 1203-1600, 1205-1589, 1207-1632, 1210-1469, 1212-1599, 1212-1602, 1214-1589, 1217-1506, 1229-1503, 1229-1508, 1237-1417, 1242-1814, 1243-1591, 1250-1592, 1253-1464, 1264-1814, 1266-1600, 1267-1599, 1275-1579, 1276-1590, 1278-1580, 1278-1599, 1278-1814, 1280-1814, 1282-1427, 1282-1552, 1282-1592, 1283-1591, 1283-1599, 1289-1586, 1290-1563, 1290-1599, 1320-1583, 1325-1591, 1337-1591, 1339-1601, 1346-1577, 1346-1592, 1432-1599, 1457-1573, 1460-1591, 1481-1582, 1481-1596, 1500-1577, 1502-1584, 1779-1814
94/7505364CB1 1705	1-273, 1-1705, 5-286, 15-299, 15-343, 25-300, 56-673, 70-321, 81-244, 84-386, 107-316, 129-677, 157-372, 298-793, 298-915, 302-587, 307-362, 327-392, 392-525, 392-566, 392-585, 392-592, 392-616, 392-629, 392-639, 392-643, 392-648, 392-669, 392-679, 392-788, 392-844, 392-895, 393-1021, 395-976, 396-514, 402-614, 407-936, 417-654, 418-689, 421-658, 423-799, 427-973, 429-1195, 431-635, 431-882, 433-632, 435-662, 440-1001, 441-1009, 442-974, 447-631, 453-745, 462-724, 462-1118, 463-704, 467-856, 468-625, 470-723, 470-1118, 480-728, 481-707, 482-741, 484-741, 484-1118, 488-1110, 494-1078, 500-1017, 515-582, 516-1008, 522-752, 533-677, 536-654, 536-803, 536-818, 539-860, 541-1207, 543-996, 545-807, 545-957, 546-1205, 548-1073, 554-819, 555-743, 556-804, 556-834, 556-843, 567-816, 568-1125, 572-852, 575-788, 575-860, 578-824, 578-1260, 579-1103, 579-1244, 580-1180, 583-764, 584-1222, 586-868, 591-902, 595-1190, 596-892, 596-1438, 597-999, 601-1204, 609-869, 610-832, 610-1207, 613-853, 617-965, 618-1112, 618-1262,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	620-887, 620-1198, 623-913, 623-1254, 624-724, 625-891, 625-920, 625-1407, 626-1161, 628-1254, 629-965, 631-1010, 634-1156, 636-923, 637-1189, 639-875, 646-1126, 649-885, 657-907, 657-1014, 658-1041, 671-1039, 672-1133, 672-1233, 674-927, 682-910, 689-1198, 690-938, 699-1331, 701-965, 701-1256, 703-968, 709-886, 709-961, 711-1056, 712-971, 712-975, 712-1255, 712-1315, 713-965, 713-1381, 714-957, 717-1372, 721-965, 721-1485, 723-1321, 727-1013, 728-1334, 730-1443, 734-1384, 735-957, 736-1164, 736-1291, 738-1392, 739-1263, 741-1159, 744-984, 745-1019, 749-1129, 750-965, 750-982, 750-989, 750-1014, 750-1143, 752-1242, 755-1033, 755-1035, 758-1449, 760-1029, 760-1094, 760-1336, 761-1205, 762-1019, 767-1021, 770-1007, 777-1013, 777-1024, 786-1282, 788-1089, 789-1115, 792-1061, 792-1102, 796-1255, 797-1185, 807-1274, 810-1056, 813-1267, 813-1407, 818-962, 818-1036, 818-1379, 819-1114, 823-965, 825-1133, 828-1013, 832-1057, 832-1245, 835-1251, 842-1075, 844-1097, 845-1031, 847-1379, 854-1087, 854-1115, 859-1280,
	862-1269, 873-1548, 877-1384, 881-1324, 882-1705, 884-1353, 892-1028, 893-1018, 894-1163, 899-1106, 901-1417, 903-1170, 915-1087, 915-1412, 915-1413, 918-1178, 918-1217, 923-1163, 926-1196, 928-1087, 937-1243, 944-1212, 946-1282, 952-1211, 953-1218, 958-1440, 967-1087, 968-1407, 971-1514, 989-1297, 991-1283, 995-1234, 995-1279, 995-1308, 995-1567, 1001-1310, 1004-1504, 1025-1353, 1026-1309, 1031-1250, 1031-1295, 1036-1270, 1037-1270, 1037-1278, 1053-1336, 1057-1292, 1057-1350, 1062-1273, 1062-1330, 1069-1316, 1069-1337, 1072-1423, 1076-1266, 1076-1354, 1081-1202, 1082-1348, 1084-1321, 1084-1328, 1084-1332, 1088-1253, 1088-1340, 1088-1397, 1089-1191, 1089-1559, 1090-1434, 1090-1481, 1091-1341, 1099-1288, 1104-1380, 1108-1356, 1108-1363, 1108-1370, 1111-1381, 1115-1379, 1116-1340, 1116-1574, 1129-1270, 1146-1551, 1148-1400, 1148-1413, 1148-1420, 1149-1565, 1151-1428, 1163-1565, 1166-1512, 1169-1369, 1169-1439, 1173-1389, 1192-1459, 1194-1433, 1196-1352, 1196-1413, 1197-1472, 1202-1435, 1208-1374, 1210-1458,
	1213-1477, 1217-1416, 1220-1477, 1228-1499, 1232-1442, 1232-1492, 1233-1475, 1234-1420, 1234-1481, 1235-1496, 1245-1505, 1256-1516, 1258-1408, 1263-1515, 1285-1544, 1314-1562, 1316-1537, 1336-1470
95/7505455CB1 1496	1-1131, 1-1140, 28-368, 28-669, 28-920, 43-595, 161-794, 167-782, 332-646, 360-1041, 409-616, 409-632, 409-913, 427-1039, 428-1030, 432-1082, 447-614, 459-1068, 467-1127, 475-1085, 497-1111, 510-1112, 515-1080, 516-948, 562-809, 571-769, 574-1095, 599-873, 651-954, 679-1135, 751-1139, 770-1125, 845-1109, 893-1496, 894-1122, 912-1105, 918-1124, 952-1124, 1065-1135
96/7505641CB1 1143	1-69, 1-197, 1-230, 1-1143, 7-137, 18-159, 25-124, 38-172, 53-513, 59-513, 230-484, 230-813, 239-453, 249-938, 262-563, 265-955, 276-1018, 278-561, 281-609, 282-803, 298-737, 312-699, 333-618, 342-749, 345-813, 347-803, 351-912, 356-609, 368-623, 377-804, 381-959, 382-845, 392-611, 400-668, 401-661, 401-933, 409-813, 410-984, 430-1016, 435-978, 439-993, 443-1018, 464-1018, 475-995, 481-993, 488-699, 490-1053, 495-1018, 519-1081, 537-688, 537-838, 537-978, 537-1002, 540-1003, 541-1003, 552-780, 560-1003, 563-1107, 572-1003, 583-1143, 594-1006, 604-833, 606-1143, 619-852, 625-933, 633-834, 633-1004, 636-914, 665-1005, 675-999, 698-936, 724-1003, 732-1003, 733-983, 739-1033, 756-1064, 779-1003, 790-1003, 800-1054, 831-999, 841-1116, 843-1098, 902-1006, 920-1123

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
97/7511242CB1 1277	1-452, 1-478, 1-592, 1-621, 1-690, 1-700, 1-709, 1-718, 1-722, 1-740, 1-749, 1-754, 1-762, 1-763, 1-775, 1-811, 1-1275, 116-649, 199-681, 199-682, 199-683, 199-685, 199-687, 199-688, 199-689, 200-688, 209-1044, 275-1111, 428-1201, 438-1275, 458-1277, 463-1277, 470-1277, 471-1277, 486-1275, 497-1275, 498-1277, 505-1277, 532-1277, 545-1277, 548-1266, 551-1277, 564-1277, 580-1275, 580-1277, 592-1277, 594-1277, 627-1273, 638-1277, 640-1277, 647-1277, 684-1275, 692-1277, 774-1277, 790-1277, 808-1275
98/7506564CB1 1812	1-242, 1-276, 1-1807, 3-229, 3-268, 6-258, 7-242, 7-265, 7-288, 9-119, 9-231, 9-277, 11-100, 11-265, 11-272, 12-289, 28-572, 54-336, 56-452, 68-580, 86-348, 238-458, 243-514, 244-533, 258-438, 258-519, 491-1020, 516-812, 604-848, 604-859, 613-1178, 621-900, 631-1131, 638-1167, 665-764, 682-858, 682-888, 682-1175, 686-1129, 695-988, 697-814, 715-937, 725-992, 731-947, 731-983, 749-1043, 755-1012, 755-1023, 755-1267, 757-1026, 770-1441, 773-1048, 777-895, 777-1003, 795-1316, 803-1066, 803-1087, 806-1055, 806-1425, 839-1100, 842-1036, 852-1467, 853-1033, 862-1563, 871-1125, 874-1120, 876-1467, 884-1127, 884-1142, 891-1114, 891-1123, 891-1183, 896-1159, 896-1189, 901-1171, 902-1465, 918-1200, 919-1194, 936-1472, 936-1517, 960-1174, 964-1197, 976-1158, 976-1472, 995-1234, 997-1221, 998-1274, 999-1267, 1006-1277, 1007-1642, 1008-1256, 1008-1293, 1027-1293, 1028-1633, 1039-1292, 1043-1630, 1048-1269, 1060-1280, 1062-1305, 1078-1631, 1078-1706, 1083-1258, 1083-1337, 1083-1350, 1105-1248, 1109-1689, 1113-1370,
	1115-1787, 1119-1252, 1119-1380, 1122-1755, 1127-1687, 1130-1412, 1132-1784, 1138-1404, 1139-1407, 1160-1744, 1166-1395, 1170-1439, 1179-1762, 1189-1476, 1196-1812, 1199-1499, 1206-1424, 1206-1762, 1207-1758, 1221-1489, 1221-1495, 1224-1497, 1224-1792, 1227-1507, 1227-1789, 1235-1520, 1237-1426, 1238-1462, 1238-1508, 1239-1812, 1240-1493, 1243-1480, 1243-1505, 1243-1796, 1243-1804, 1245-1786, 1245-1789, 1248-1487, 1268-1762, 1287-1467, 1287-1500, 1290-1812, 1296-1799, 1298-1717, 1307-1517, 1327-1812, 1330-1477, 1330-1587, 1331-1577, 1339-1798, 1341-1549, 1341-1574, 1350-1577, 1353-1505, 1354-1584, 1366-1798, 1370-1596, 1372-1625, 1373-1812, 1386-1620, 1388-1812, 1401-1658, 1401-1743, 1401-1798, 1409-1619, 1409-1812, 1421-1812, 1423-1640, 1434-1799, 1447-1721, 1467-1758, 1468-1671, 1469-1683, 1470-1720, 1488-1766, 1495-1707, 1495-1779, 1495-1793, 1499-1812, 1504-1734, 1518-1806, 1540-1705, 1543-1799, 1543-1812, 1558-1778, 1620-1798, 1624-1812, 1626-1812, 1627-1812, 1675-1812, 1693-1812
99/7509076CB1 1398	1-237, 1-1398, 212-851, 228-1027, 231-934, 239-509, 245-901, 252-747, 273-603, 291-674, 336-772, 440-964, 588-852, 606-830, 606-1336, 607-1301, 612-1372, 616-935, 683-1332, 803-1013, 831-1398, 840-1398, 850-1392, 865-1395, 866-1342, 868-1398, 871-1366, 876-1398, 877-1245, 884-1398, 899-1315, 899-1398, 906-1398, 924-1395, 928-1254, 930-1391, 938-1398, 940-1392, 940-1398, 943-1398, 953-1394, 957-1398, 965-1398, 968-1388, 969-1227, 973-1388, 987-1388, 987-1395, 997-1333, 1004-1349, 1004-1387, 1021-1270, 1024-1398, 1030-1388, 1032-1329, 1032-1388, 1040-1198, 1044-1391, 1044-1398, 1045-1301, 1054-1389, 1056-1395, 1057-1398, 1063-1388, 1071-1140, 1095-1350, 1116-1291, 1116-1384, 1136-1391, 1142-1390, 1143-1397, 1152-1388, 1158-1393, 1190-1387, 1190-1388, 1197-1327, 1208-1398, 1209-1388, 1213-1381, 1213-1391, 1227-1388, 1239-1398, 1249-1390, 1256-1398, 1271-1388

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
100/7506666CB1 1107	1-243, 1-330, 1-460, 1-471, 1-532, 1-544, 1-548, 1-570, 1-597, 1-606, 1-621, 1-629, 1-1107, 15-703, 16-667, 16-733, 16-862, 17-922, 45-532, 45-616, 45-625, 45-713, 45-726, 45-752, 45-780, 45-921, 48-742, 51-740, 69-478, 126-731, 138-923, 262-638, 267-667, 295-687, 317-541, 327-770, 335-514, 336-737, 345-739, 352-571, 356-503, 361-503, 361-804, 431-1107, 442-695, 481-680, 487-676, 493-648, 533-790, 534-788, 557-734, 618-893, 656-1092, 687-807, 699-1092, 740-848, 952-1091
101/7511731CB1 991	1-141, 4-225, 4-236, 4-284, 5-235, 5-249, 5-975, 6-141, 6-328, 7-215, 8-231, 8-241, 9-147, 9-167, 17-273, 18-244, 41-241, 42-110, 48-228, 48-231, 69-250, 125-160, 125-162, 125-164, 125-165, 125-169, 125-172, 125-173, 125-177, 125-178, 125-185, 125-192, 125-194, 125-199, 125-200, 125-204, 125-207, 125-213, 125-214, 125-215, 125-218, 125-221, 125-223, 125-224, 125-225, 125-227, 125-234, 125-236, 125-240, 125-246, 125-248, 125-250, 125-252, 125-260, 125-264, 125-267, 125-274, 125-275, 125-277, 125-278, 125-280, 125-284, 125-285, 125-287, 125-289, 125-291, 125-292, 125-302, 125-308, 125-315, 125-316, 125-317, 125-323, 125-327, 125-330, 125-337, 125-338, 125-343, 125-344, 125-346, 125-347, 125-348, 125-359, 125-363, 125-371, 125-378, 125-386, 125-411, 125-436, 125-473, 125-564, 125-622, 125-659, 126-298, 126-378, 127-400, 127-401, 128-233, 128-315, 128-346, 131-436, 134-350, 135-407, 136-348, 136-378, 137-227, 137-369, 139-315, 139-341, 139-406, 139-446, 145-315, 146-259, 146-380, 146-409, 147-315, 148-315,
	155-378, 155-393, 156-250, 161-413, 175-378, 182-250, 182-390, 182-409, 182-410, 182-413, 182-422, 182-426, 182-436, 184-315, 184-413, 184-469, 185-378, 185-438, 188-226, 188-227, 188-275, 188-285, 188-287, 188-336, 188-391, 188-413, 188-722, 189-346, 189-436, 190-460, 190-461, 191-413, 196-473, 197-411, 198-250, 198-467, 199-420, 199-434, 200-315, 200-432, 202-250, 202-378, 202-404, 202-502, 202-509, 206-413, 207-471, 208-322, 209-378, 210-378, 211-378, 218-436, 223-250, 224-250, 224-478, 240-378, 242-477, 243-534, 244-315, 244-436, 244-450, 244-453, 244-469, 244-473, 245-436, 245-473, 248-338, 248-343, 248-346, 248-348, 248-358, 248-359, 248-364, 248-367, 248-371, 248-399, 248-404, 248-483, 248-745, 249-413, 249-473, 250-523, 250-524, 251-475, 256-564, 257-469, 258-530, 259-473, 259-501, 260-378, 260-473, 262-436, 262-464, 262-529, 262-572, 266-503, 267-532, 268-385, 269-436, 270-436, 278-473, 278-516, 285-378, 288-532, 301-378, 301-513, 301-533, 301-534, 301-545, 301-549, 301-564, 305-473, 307-436,
	307-532, 307-597, 308-473, 308-557, 311-353, 311-401, 311-409, 311-411, 311-414, 311-421, 311-422, 311-427, 311-430, 311-435, 311-459, 311-464, 311-514, 311-532, 311-750, 312-483, 312-564, 313-586, 313-587, 314-532, 319-622, 320-542, 321-378, 321-593, 322-543, 322-564, 323-434, 323-555, 325-378, 325-473, 325-527, 325-592, 325-632, 329-566, 330-595, 331-445, 332-473, 333-473, 334-473, 341-564, 341-579, 342-436, 358-564, 363-599, 368-436, 368-576, 368-595, 368-596, 368-599, 368-608, 368-612, 368-622, 370-473, 370-599, 370-655, 371-564, 371-624, 374-412, 374-413, 374-461, 374-466, 374-469, 374-473, 374-522, 374-577, 374-599, 374-722, 375-532, 375-622, 376-646, 376-647, 377-599, 382-659, 383-605, 384-436, 384-653, 385-606, 385-620, 386-501, 386-618, 388-436, 388-564, 388-590, 388-688, 388-695, 392-599, 393-657, 394-508, 395-564, 396-564, 397-564, 404-622, 410-664, 426-564, 428-663, 429-720, 430-473, 430-622, 430-636, 430-639, 430-655, 430-659, 431-622, 431-659, 433-983, 434-474, 434-524, 434-529, 434-532,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	434-534, 434-544, 434-545, 434-550, 434-553, 434-557, 434-585, 434-590, 434-674, 434-745, 435-599, 435-659, 436-706, 436-710, 437-661, 442-914, 443-655, 444-716, 445-659, 445-687, 446-564, 446-659, 448-622, 448-650, 448-725, 448-745, 452-689, 453-718, 454-571, 464-702, 465-564, 474-718, 486-699, 486-724, 487-564, 487-719, 487-731, 487-735, 491-659, 493-622, 493-722, 494-659, 494-745, 497-539, 497-587, 497-597, 497-600, 497-603, 497-607, 497-608, 497-613, 497-616, 497-621, 497-622, 497-645, 497-650, 497-700, 497-722, 497-750, 498-670, 498-767, 498-824, 498-914, 498-932, 498-983, 498-990, 499-722, 499-745, 500-722, 505-722, 506-728, 507-564, 507-745, 508-729, 508-750, 509-599, 509-741, 511-564, 511-659, 511-712, 511-731, 511-750, 515-722, 516-745, 517-631, 518-659, 519-659, 520-659, 527-722, 527-792, 538-745, 544-801, 554-622, 554-722, 556-659, 556-745, 557-750, 557-808, 560-598, 560-599, 560-647, 560-652, 560-655, 560-659, 560-706, 560-745, 560-750, 561-722, 562-745, 562-750, 563-745, 568-745, 569-722,
	570-622, 570-750, 571-722, 572-687, 572-722, 574-622, 574-722, 574-750, 574-813, 578-745, 579-750, 580-694, 581-991, 582-826, 583-810, 590-722, 595-622, 596-622, 596-750, 612-769, 614-750, 615-750, 616-659, 616-722, 616-745, 617-722, 617-750, 619-740, 620-660, 620-710, 620-718, 620-720, 620-722, 620-726, 620-730, 620-731, 620-736, 620-739, 620-744, 620-745, 620-750, 621-745, 622-750, 623-750, 628-743, 629-744, 630-722, 631-745, 632-745, 632-750, 634-722, 634-745, 638-742, 639-750, 640-750, 650-750, 651-868, 660-750, 673-750, 673-902, 677-745, 679-722, 679-750, 680-745, 683-722, 683-741, 683-750, 684-750, 685-722, 685-745, 686-750, 693-745, 693-914, 694-750, 695-745, 695-750, 697-745, 697-750, 697-915, 697-945, 698-745, 702-745, 708-745, 714-743, 718-933, 719-951, 719-957, 735-983, 785-975
102/7511735CB1 863	1-226, 1-237, 1-239, 1-242, 1-245, 1-252, 1-268, 1-845, 2-137, 2-233, 3-226, 5-230, 7-173, 38-106, 100-366, 121-158, 121-160, 121-167, 121-170, 121-174, 121-176, 121-182, 121-188, 121-189, 121-193, 121-200, 121-205, 121-212, 121-219, 121-220, 121-221, 121-230, 121-242, 121-244, 121-251, 121-263, 121-268, 121-282, 121-283, 121-286, 121-292, 121-293, 121-305, 121-309, 121-316, 121-323, 121-328, 121-336, 121-343, 121-434, 121-492, 121-529, 122-286, 122-343, 124-229, 124-309, 129-388, 129-427, 130-343, 131-188, 132-343, 133-343, 135-188, 135-309, 143-309, 151-343, 168-343, 169-275, 172-345, 176-383, 176-402, 176-403, 176-404, 176-434, 181-343, 184-236, 184-284, 184-286, 184-293, 184-296, 184-300, 184-302, 184-305, 184-307, 184-328, 184-403, 184-592, 185-353, 185-434, 190-499, 192-451, 193-412, 194-251, 195-413, 196-425, 198-251, 198-343, 206-343, 214-434, 231-434, 232-335, 235-403, 241-446, 241-465, 241-466, 241-469, 241-492, 244-434, 247-299, 247-343, 247-372, 247-388, 247-466, 247-615, 248-402, 248-492,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	255-511, 256-475, 257-309, 258-476, 259-488, 261-309, 261-434, 269-434, 277-492, 303-395, 303-466, 303-492, 303-506, 303-525, 303-529, 303-562, 304-492, 307-409, 307-413, 307-416, 307-420, 307-422, 307-425, 307-428, 307-435, 307-451, 307-552, 307-620, 308-469, 308-529, 313-615, 315-574, 316-525, 318-529, 319-529, 321-492, 327-458, 344-588, 352-529, 355-522, 357-569, 357-589, 357-594, 357-859, 363-492, 364-529, 367-419, 367-467, 367-469, 367-476, 367-479, 367-483, 367-485, 367-488, 367-490, 367-511, 367-589, 368-540, 368-637, 368-694, 368-784, 368-802, 368-859, 368-863, 373-592, 375-589, 376-598, 377-434, 378-599, 379-611, 381-434, 381-529, 389-529, 397-662, 414-671, 415-518, 418-589, 424-592, 427-678, 430-482, 430-525, 430-530, 430-558, 430-574, 430-615, 431-592, 438-615, 439-592, 440-492, 441-592, 442-592, 444-492, 444-683, 452-696, 460-592, 486-581, 486-592, 486-615, 487-592, 490-524, 490-590, 490-592, 490-599, 490-602, 490-606, 490-608, 490-611, 490-613, 490-620, 491-615, 496-612, 498-620, 499-614,
	501-615, 502-615, 504-592, 510-620, 530-620, 541-620, 543-610, 543-620, 547-615, 549-592, 550-615, 553-620, 554-620, 561-620, 563-784, 564-620, 565-620, 567-615, 567-785, 578-615, 584-613, 589-827, 613-738, 655-845
103/7511729CB1 807	1-233, 1-803, 7-240, 24-305, 26-256, 26-348, 27-302, 29-233, 31-232, 31-236, 31-321, 33-265, 33-273, 33-279, 33-288, 34-301, 35-217, 35-260, 35-282, 35-285, 35-289, 35-305, 35-309, 35-316, 35-323, 36-244, 36-282, 36-297, 36-298, 36-347, 37-265, 37-322, 38-250, 38-273, 38-282, 38-287, 38-289, 38-425, 38-696, 39-266, 39-271, 40-312, 40-322, 40-334, 41-328, 41-344, 42-303, 44-193, 44-284, 45-349, 46-321, 48-260, 49-201, 49-218, 49-359, 50-187, 50-274, 51-313, 52-317, 53-302, 53-343, 56-159, 56-356, 56-485, 58-270, 58-303, 58-308, 58-322, 58-327, 58-330, 58-332, 60-340, 63-270, 64-281, 64-309, 65-337, 67-341, 67-348, 68-266, 68-336, 69-274, 71-341, 72-279, 72-334, 73-326, 73-358, 74-306, 74-345, 86-358, 95-340, 97-368, 106-410, 109-371, 120-353, 129-405, 135-365, 161-443, 180-440, 180-461, 181-467, 185-413, 185-433, 189-470, 200-461, 208-456, 210-472, 212-378, 212-411, 212-446, 229-460, 233-485, 295-354, 298-728, 360-475, 438-485, 445-485, 478-653, 478-699, 478-802, 478-807, 491-735, 492-734, 497-807, 504-803,
	507-807, 513-769, 513-793, 513-803, 514-746, 515-806, 522-789, 523-766, 523-803, 524-803, 530-778, 530-797, 530-807, 533-765, 536-802, 536-803, 544-797, 546-803, 548-803, 551-804, 553-804, 554-803, 556-803, 558-803, 565-803, 573-803, 586-725, 586-746, 586-800, 596-800, 598-804, 598-807, 601-807, 603-691, 609-733, 610-797, 613-681, 626-797, 647-807, 663-803, 664-803, 664-807, 666-797, 669-803, 684-803

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1900-2159, 1917-2161, 1917-2244, 1917-2383, 1918-2136, 1920-2453, 1934-2324, 1948-2396, 1958-2453, 1965-2435, 1970-2444, 1976-2452, 1976-2453, 1983-2353, 1983-2423, 1983-2447, 1984-2152, 1996-2453, 1998-2247, 1998-2442, 1999-2446, 2000-2451, 2001-2405, 2005-2287, 2009-2453, 2011-2439, 2018-2448, 2018-2449, 2019-2444, 2020-2199, 2020-2444, 2023-2443, 2028-2296, 2029-2199, 2034-2444, 2034-2453, 2037-2445, 2039-2444, 2040-2306, 2045-2453, 2046-2348, 2059-2447, 2060-2444, 2063-2444, 2063-2453, 2068-2295, 2070-2444, 2072-2444, 2073-2401, 2077-2215, 2078-2444, 2080-2344, 2080-2444, 2082-2444, 2083-2444, 2084-2273, 2091-2419, 2099-2453, 2107-2381, 2110-2448, 2111-2450, 2113-2373, 2118-2453, 2122-2346, 2124-2444, 2136-2444, 2147-2449, 2151-2452, 2155-2382, 2166-2424, 2174-2443, 2193-2447, 2193-2453, 2194-2446, 2204-2440, 2214-2448, 2220-2396, 2220-2452, 2225-2380, 2225-2444, 2226-2444, 2227-2395, 2228-2444, 2236-2444, 2238-2444, 2264-2453, 2278-2443, 2285-2453, 2299-2444, 2315-2450, 2338-2444
105/7511628CB1 1630	1-1589, 20-428, 115-480, 169-748, 235-509, 271-571, 293-493, 293-503, 293-524, 293-537, 293-742, 293-765, 293-790, 293-842, 293-845, 293-851, 293-892, 295-893, 308-665, 320-905, 332-645, 341-811, 346-657, 362-895, 365-574, 376-602, 384-629, 393-1117, 394-610, 394-618, 394-835, 405-600, 405-636, 413-1135, 431-689, 445-666, 445-692, 445-979, 461-757, 461-1109, 462-739, 469-1107, 481-725, 481-773, 483-740, 548-810, 553-1107, 553-1108, 561-1122, 567-1230, 569-1135, 571-1134, 572-1190, 575-1093, 603-872, 615-1111, 620-833, 620-1205, 623-845, 630-1204, 644-1082, 653-879, 660-909, 660-1095, 671-895, 672-1138, 677-949, 686-1563, 689-1219, 690-1147, 696-959, 696-1270, 698-981, 699-1325, 705-933, 707-1194, 707-1196, 716-1351, 732-903, 735-974, 747-1027, 755-1347, 758-1022, 766-970, 767-1356, 769-984, 769-1170, 776-1016, 779-1457, 792-1054, 792-1084, 793-1039, 793-1042, 798-1390, 804-1094, 804-1294, 811-1086, 813-1046, 818-1442, 821-1085, 829-1587, 830-1092, 844-1420, 845-1515, 857-1092, 857-1347, 863-1518, 872-1102,
	872-1460, 876-1393, 892-1320, 896-1162, 898-1588, 899-1156, 899-1307, 899-1415, 915-1575, 924-1530, 924-1532, 925-1535, 929-1494, 941-1588, 942-1211, 943-1469, 945-1588, 968-1587, 972-1219, 975-1568, 975-1575, 979-1588, 980-1102, 993-1238, 993-1242, 998-1236, 1009-1432, 1013-1587, 1017-1592, 1019-1285, 1022-1102, 1033-1452, 1037-1575, 1037-1583, 1052-1542, 1058-1592, 1059-1284, 1068-1587, 1081-1259, 1081-1264, 1081-1296, 1081-1332, 1081-1381, 1089-1295, 1090-1593, 1092-1594, 1099-1298, 1101-1592, 1110-1576, 1112-1355, 1114-1575, 1123-1575, 1126-1394, 1126-1529, 1126-1530, 1126-1574, 1126-1575, 1133-1589, 1133-1596, 1135-1575, 1138-1573, 1145-1575, 1150-1588, 1155-1630, 1158-1575, 1160-1531, 1162-1397, 1165-1576, 1169-1561, 1172-1575, 1172-1595, 1175-1599, 1190-1443, 1191-1442, 1196-1608, 1200-1523, 1202-1457, 1205-1476, 1210-1572, 1210-1579, 1212-1574, 1213-1438, 1213-1575, 1215-1574, 1220-1575, 1220-1577, 1221-1574, 1224-1575, 1228-1587, 1234-1575, 1238-1437, 1239-1580, 1245-1575, 1245-1593, 1248-1575,
	1252-1575, 1264-1516, 1271-1575, 1281-1590, 1286-1575, 1312-1594, 1323-1590, 1328-1574, 1334-1590, 1336-1590, 1337-1552, 1337-1575, 1338-1574, 1355-1575, 1359-1562, 1367-1590, 1378-1575, 1381-1574, 1387-1584, 1400-1619, 1447-1624, 1469-1604, 1488-1607, 1490-1593, 1511-1624

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
106/7511657CB1 1890	1-95, 2-204, 2-281, 2-1876, 5-144, 6-208, 6-278, 8-70, 8-266, 8-285, 11-257, 21-296, 22-286, 111-716, 320-572, 322-597, 322-599, 324-460, 342-487, 342-522, 342-582, 342-623, 349-471, 354-604, 362-654, 368-559, 368-596, 370-623, 375-621, 394-632, 396-648, 397-636, 398-637, 398-685, 399-672, 409-620, 416-668, 427-662, 427-1174, 428-654, 437-578, 455-719, 466-613, 466-692, 468-715, 507-1018, 514-647, 519-754, 532-1122, 532-1174, 533-584, 540-664, 540-689, 540-1174, 574-742, 578-794, 588-834, 588-903, 591-832, 592-738, 592-866, 597-841, 620-648, 620-890, 637-1170, 639-840, 639-861, 643-668, 655-1216, 660-892, 663-906, 663-1005, 689-933, 739-1174, 743-1001, 743-1016, 747-1022, 747-1274, 751-988, 751-1080, 752-1029, 753-974, 758-1023, 758-1030, 764-975, 777-1023, 777-1045, 777-1056, 777-1064, 778-1029, 778-1048, 787-1037, 788-1105, 792-944, 794-1049, 795-1046, 799-1037, 799-1204, 802-1022, 802-1037, 802-1115, 805-1048, 810-1087, 813-1067, 815-1056, 819-1063, 819-1155, 820-1135, 821-1003, 821-1049, 821-1056,
	822-1073, 822-1148, 823-1048, 823-1064, 823-1065, 823-1074, 823-1081, 823-1087, 823-1088, 823-1122, 823-1144, 824-1059, 824-1096, 825-1040, 825-1069, 825-1071, 825-1084, 825-1087, 825-1102, 826-1053, 826-1067, 826-1076, 826-1086, 826-1088, 826-1097, 826-1143, 827-1087, 827-1185, 828-947, 828-1049, 828-1056, 828-1058, 828-1060, 828-1067, 828-1069, 828-1070, 828-1074, 828-1083, 828-1095, 828-1104, 828-1109, 828-1116, 828-1162, 828-1174, 828-1190, 830-1077, 830-1080, 830-1099, 831-1031, 831-1044, 831-1047, 831-1068, 831-1102, 831-1131, 833-1114, 833-1164, 834-1109, 835-1055, 835-1101, 835-1106, 836-1353, 838-1080, 838-1133, 839-1077, 839-1099, 840-1100, 840-1101, 840-1109, 841-1167, 844-1092, 847-1044, 847-1064, 847-1083, 847-1085, 847-1123, 849-1052, 850-1109, 850-1159, 851-1082, 851-1105, 851-1108, 851-1127, 852-1127, 853-1032, 853-1095, 853-1130, 854-1115, 855-1116, 855-1132, 857-1106, 857-1140, 859-1094, 862-1105, 862-1174, 866-1135, 868-1013, 868-1106, 871-1101, 871-1106, 871-1142, 872-1112, 873-1174,
	874-1137, 877-963, 878-1075, 878-1108, 878-1141, 881-1193, 885-1124, 888-1208, 894-1122, 901-1140, 902-1132, 902-1197, 904-1162, 905-1195, 907-1125, 907-1127, 918-1134, 920-1103, 920-1191, 925-1708, 926-1145, 927-1219, 927-1547, 931-1220, 940-1196, 944-1221, 948-1207, 957-1189, 977-1193, 978-1236, 979-1767, 981-1213, 986-1214, 991-1824, 993-1800, 996-1212, 999-1203, 1008-1100, 1017-1800, 1030-1301, 1032-1556, 1042-1331, 1064-1680, 1068-1310, 1077-1325, 1078-1811, 1113-1773, 1114-1174, 1116-1314, 1116-1401, 1117-1378, 1117-1781, 1127-1382, 1129-1392, 1129-1416, 1130-1356, 1133-1353, 1133-1365, 1133-1373, 1133-1389, 1133-1395, 1133-1396, 1133-1425, 1137-1427, 1138-1362, 1140-1402, 1140-1433, 1141-1347, 1141-1355, 1142-1373, 1144-1371, 1145-1386, 1145-1569, 1146-1404, 1147-1366, 1150-1773, 1155-1430, 1156-1374, 1156-1440, 1160-1799, 1160-1874, 1161-1408, 1161-1817, 1161-1829, 1161-1830, 1161-1831, 1161-1832, 1161-1834, 1161-1852, 1162-1765, 1162-1834, 1164-1388, 1164-1422, 1164-1433, 1166-1397, 1166-1428,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
106 - Cont'd	1167-1877, 1170-1380, 1170-1396, 1172-1379, 1172-1396, 1172-1410, 1172-1435, 1173-1401, 1173-1431, 1173-1432, 1173-1793, 1176-1393, 1177-1425, 1182-1352, 1182-1462, 1183-1412, 1184-1432, 1184-1811, 1185-1390, 1185-1392, 1187-1409, 1187-1877, 1189-1797, 1189-1884, 1190-1680, 1191-1833, 1191-1876, 1192-1876, 1194-1394, 1194-1441, 1195-1772, 1197-1477, 1197-1518, 1200-1450, 1200-1835, 1203-1469, 1204-1858, 1205-1404, 1206-1450, 1207-1473, 1207-1492, 1208-1831, 1209-1768, 1209-1808, 1210-1877, 1214-1479, 1215-1680, 1217-1462, 1220-1493, 1220-1819, 1227-1438, 1227-1458, 1227-1469, 1227-1484, 1239-1849, 1240-1478, 1241-1511, 1242-1806, 1242-1875, 1243-1849, 1245-1850, 1247-1519, 1250-1421, 1250-1503, 1250-1552, 1260-1522, 1262-1472, 1262-1576, 1262-1709, 1263-1539, 1265-1536, 1265-1822, 1265-1884, 1266-1890, 1269-1526, 1281-1519, 1284-1556, 1286-1559, 1287-1598, 1289-1498, 1289-1558, 1289-1577, 1290-1805, 1293-1546, 1296-1768, 1301-1836, 1308-1800, 1313-1890, 1319-1595, 1321-1777, 1322-1619, 1325-1599, 1326-1589,
	1328-1575, 1338-1655, 1338-1813, 1342-1557, 1342-1572, 1342-1611, 1342-1650, 1343-1630, 1344-1540, 1344-1579, 1346-1598, 1350-1582, 1350-1650, 1350-1676, 1351-1628, 1351-1663, 1356-1622, 1359-1622, 1363-1470, 1363-1587, 1363-1616, 1364-1715, 1365-1587, 1368-1629, 1368-1631, 1368-1647, 1369-1663, 1371-1581, 1371-1594, 1371-1602, 1371-1615, 1371-1628, 1372-1529, 1372-1590, 1372-1625, 1373-1607, 1373-1617, 1373-1625, 1375-1612, 1376-1584, 1376-1607, 1376-1649, 1376-1744, 1376-1860, 1376-1871, 1377-1589, 1377-1631, 1377-1642, 1377-1651, 1378-1615, 1378-1744, 1384-1698, 1385-1617, 1386-1877, 1393-1633, 1393-1655, 1395-1605, 1395-1640, 1398-1625, 1398-1637, 1398-1877, 1399-1663, 1401-1606, 1401-1652, 1402-1637, 1402-1678, 1403-1639, 1404-1546, 1404-1637, 1407-1641, 1411-1686, 1413-1598, 1414-1672, 1414-1676, 1414-1875, 1416-1631, 1417-1610, 1418-1658, 1418-1766, 1420-1650, 1421-1876, 1423-1698, 1423-1838, 1425-1657, 1425-1876, 1426-1695, 1426-1880, 1428-1755, 1428-1880, 1429-1875, 1430-1654, 1430-1706, 1434-1643,
	1434-1648, 1434-1666, 1434-1678, 1434-1876, 1437-1697, 1437-1738, 1437-1831, 1437-1880, 1438-1772, 1439-1713, 1440-1667, 1440-1876, 1441-1669, 1442-1726, 1444-1673, 1444-1687, 1446-1653, 1446-1663, 1446-1687, 1449-1886, 1451-1681, 1451-1875, 1454-1714, 1455-1697, 1455-1729, 1456-1875, 1461-1659, 1461-1710, 1461-1713, 1462-1703, 1466-1712, 1466-1821, 1470-1777, 1470-1882, 1478-1712, 1482-1872, 1484-1706, 1491-1697, 1492-1733, 1492-1738, 1492-1877, 1494-1850, 1494-1875, 1495-1712, 1498-1875, 1498-1887, 1499-1774, 1503-1862, 1505-1875, 1506-1875, 1507-1771, 1511-1750, 1515-1875, 1521-1766, 1522-1793, 1526-1875, 1526-1877, 1530-1766, 1532-1760, 1532-1876, 1535-1881, 1536-1758, 1536-1880, 1540-1759, 1542-1707, 1544-1792, 1550-1791, 1551-1751, 1551-1832, 1552-1829, 1557-1801, 1558-1783, 1558-1797, 1558-1819, 1559-1890, 1560-1804, 1563-1808, 1567-1886, 1567-1890, 1570-1766, 1570-1780, 1571-1876, 1571-1877, 1574-1880, 1577-1820, 1581-1821, 1584-1875, 1585-1801, 1585-1839, 1585-1843, 1585-1875, 1588-1767, 1588-1815,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1588-1818, 1588-1822, 1588-1835, 1588-1875, 1588-1882, 1589-1861, 1590-1803, 1590-1832, 1590-1854, 1590-1875, 1597-1833, 1598-1877, 1599-1876, 1600-1874, 1602-1812, 1602-1842, 1602-1877, 1605-1856, 1610-1816, 1610-1824, 1610-1839, 1610-1846, 1610-1875, 1612-1846, 1615-1841, 1615-1856, 1615-1874, 1619-1875, 1625-1873, 1627-1841, 1628-1877, 1629-1841, 1631-1829, 1631-1841, 1631-1869, 1631-1878, 1632-1847, 1633-1854, 1638-1880, 1640-1878, 1640-1885, 1642-1820, 1643-1867, 1643-1879, 1644-1879, 1646-1849, 1646-1884, 1647-1890, 1648-1890, 1651-1876, 1652-1836, 1653-1875, 1655-1774, 1658-1847, 1661-1875, 1662-1880, 1665-1790, 1665-1890, 1667-1863, 1667-1875, 1668-1877, 1669-1785, 1674-1877, 1675-1886, 1679-1876, 1681-1860, 1683-1876, 1684-1875, 1685-1881, 1691-1877, 1691-1879, 1692-1877, 1693-1780, 1693-1827, 1693-1872, 1693-1875, 1694-1875, 1694-1890, 1696-1886, 1697-1877, 1700-1875, 1702-1876, 1707-1876, 1709-1877, 1710-1877, 1710-1890, 1711-1881, 1711-1890, 1714-1879, 1715-1875, 1716-1880, 1717-1875, 1718-1875,
	1721-1877, 1722-1877, 1723-1848, 1723-1876, 1727-1877, 1729-1876, 1730-1837, 1730-1879, 1740-1875, 1748-1861, 1753-1877, 1756-1868, 1769-1880, 1771-1877, 1779-1877, 1790-1877, 1800-1875, 1807-1877, 1808-1867, 1808-1875, 1809-1876, 1826-1875
107/7512343CB1 1143	1-883, 55-1143, 364-1143
108/7512357CB1 1739	1-885, 1-1710, 89-976, 184-997, 815-1739, 817-1658, 819-1512
109/7511046CB1 837	1-137, 1-360, 1-382, 1-434, 1-445, 1-536, 1-637, 1-763, 3-304, 136-833, 208-833, 258-828, 258-829, 263-649, 383-814, 460-835, 466-837, 468-837, 517-829, 601-833
110/7512332CB1 736	1-674, 1-736, 253-454, 288-338, 295-386, 316-505, 358-403, 402-452, 460-505, 460-551, 461-505
111/7511219CB1 412	1-288, 1-412

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
112/7510933CB1 2815	1-253, 10-280, 20-252, 20-2815, 30-288, 36-263, 43-359, 43-595, 48-295, 50-186, 50-315, 56-326, 56-343, 56-353, 57-150, 57-196, 57-285, 57-346, 58-148, 58-287, 58-321, 59-328, 59-366, 62-274, 66-312, 66-345, 73-311, 101-308, 123-402, 138-333, 140-394, 258-402, 396-1011, 396-1016, 447-1277, 561-1237, 578-1020, 622-720, 679-1250, 679-1291, 701-1206, 820-1535, 834-1603, 835-1091, 837-1133, 846-1440, 848-1455, 872-1238, 879-1004, 886-1416, 890-1478, 894-1428, 898-1665, 907-1194, 912-1322, 919-1493, 921-1578, 922-1511, 924-1706, 929-1415, 933-1744, 937-1433, 943-1596, 946-1191, 954-1199, 959-1183, 959-1199, 959-1233, 959-1545, 964-1206, 980-1623, 996-1199, 1002-1706, 1011-1706, 1015-1601, 1019-1204, 1029-1740, 1046-1706, 1047-1661, 1051-1699, 1058-1706, 1065-1389, 1069-1613, 1069-1706, 1072-1731, 1080-1702, 1090-1801, 1106-1345, 1106-1417, 1106-1634, 1109-1347, 1120-1338, 1120-1880, 1121-1701, 1127-1870, 1129-1706, 1139-1927, 1149-1729, 1159-1426, 1163-1927, 1164-1724, 1169-1796, 1176-1270, 1184-1850, 1219-1457,
	1219-1692, 1221-1697, 1227-1411, 1230-1843, 1230-1947, 1238-1969, 1252-1852, 1263-1958, 1265-1511, 1270-1535, 1271-1697, 1298-1534, 1306-1537, 1306-1561, 1306-1782, 1316-1581, 1321-1656, 1338-1706, 1342-1850, 1343-2244, 1351-1593, 1378-1708, 1378-2071, 1383-1635, 1408-1922, 1415-1896, 1416-1711, 1420-2052, 1426-2065, 1429-1961, 1430-1904, 1449-2037, 1451-1753, 1453-1702, 1453-1744, 1455-1705, 1455-1933, 1458-2141, 1462-1739, 1462-2003, 1477-2139, 1496-2142, 1501-2071, 1520-2101, 1520-2109, 1522-2020, 1522-2085, 1524-2225, 1538-1765, 1548-1874, 1553-2185, 1553-2213, 1557-1826, 1564-1959, 1568-2339, 1579-2367, 1589-1845, 1593-1870, 1595-2231, 1596-1893, 1623-2149, 1623-2402, 1646-1852, 1646-2303, 1650-2303, 1654-2276, 1683-1952, 1697-2015, 1703-2020, 1707-2391, 1723-2244, 1733-2022, 1741-2355, 1741-2358, 1745-2612, 1747-2016, 1750-2378, 1755-2358, 1757-2340, 1758-2067, 1761-2009, 1762-2284, 1765-1975, 1765-2010, 1765-2171, 1765-2365, 1767-2050, 1767-2074, 1771-2007, 1771-2014, 1771-2193, 1777-2019, 1777-2389,
	1783-2070, 1794-2324, 1795-2421, 1799-2442, 1805-2093, 1805-2097, 1812-2458, 1816-2437, 1823-2347, 1826-2095, 1828-2078, 1848-2538, 1850-1997, 1852-2156, 1855-2141, 1862-2463, 1863-2374, 1865-2164, 1875-2769, 1898-2624, 1904-2396, 1908-2426, 1909-2196, 1914-2192, 1917-2184, 1921-2198, 1929-2448, 1933-2637, 1939-2381, 1940-2167, 1942-2171, 1943-2537, 1946-2558, 1947-2516, 1947-2529, 1948-2334, 1949-2760, 1951-2474, 1951-2576, 1952-2073, 1959-2523, 1963-2755, 1971-2504, 1980-2676, 1990-2670, 1999-2238, 2007-2557, 2029-2253, 2029-2280, 2038-2422, 2059-2344, 2074-2505, 2074-2552, 2082-2716, 2083-2735, 2086-2667, 2087-2352, 2088-2512, 2089-2339, 2092-2279, 2094-2710, 2096-2374, 2099-2431, 2102-2289, 2102-2370, 2107-2680, 2113-2790, 2124-2398, 2130-2670, 2134-2815, 2140-2668, 2141-2801, 2179-2815, 2181-2433, 2183-2469, 2183-2764, 2198-2767, 2199-2783, 2207-2388, 2207-2773, 2211-2810, 2214-2802, 2215-2767, 2218-2488, 2218-2760, 2220-2449, 2220-2804, 2230-2693, 2231-2432, 2231-2515, 2233-2507, 2233-2787, 2233-2815,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	2242-2769, 2256-2487, 2256-2807, 2260-2741, 2261-2513, 2266-2525, 2283-2527, 2283-2610, 2283-2749, 2284-2502, 2286-2815, 2300-2690, 2314-2762, 2324-2806, 2324-2810, 2331-2801, 2336-2810, 2342-2815, 2349-2719, 2349-2789, 2349-2813, 2350-2518, 2362-2815, 2364-2613, 2364-2808, 2365-2812, 2366-2815, 2367-2771, 2371-2653, 2375-2815, 2377-2805, 2384-2814, 2384-2815, 2385-2810, 2386-2565, 2386-2810, 2389-2809, 2394-2662, 2395-2565, 2400-2810, 2400-2815, 2403-2811, 2405-2810, 2406-2672, 2411-2815, 2412-2714, 2425-2813, 2426-2810, 2429-2810, 2429-2815, 2434-2661, 2436-2810, 2438-2810, 2439-2767, 2443-2581, 2444-2810, 2446-2710, 2446-2810, 2448-2810, 2449-2810, 2450-2639, 2457-2785, 2465-2815, 2473-2747, 2476-2814, 2477-2815, 2479-2739, 2484-2815, 2488-2712, 2490-2810, 2502-2810, 2513-2815, 2517-2810, 2521-2748, 2532-2790, 2540-2809, 2559-2813, 2559-2815, 2560-2812, 2570-2806, 2580-2814, 2586-2762, 2586-2810, 2591-2746, 2591-2810, 2592-2810, 2593-2761, 2594-2810, 2602-2810, 2604-2810, 2630-2815, 2644-2809, 2651-2815, 2665-2810, 2681-2815, 2704-2810
113/7511461CB1 723	1-707, 79-631, 145-399, 145-413, 216-723, 228-723, 267-707, 272-723, 392-632, 403-707
114/7511808CB1 2748	1-154, 1-197, 2-2622, 5-166, 10-197, 23-859, 24-178, 24-191, 27-193, 29-289, 198-744, 203-514, 220-659, 220-748, 220-759, 221-544, 234-514, 237-774, 239-479, 245-859, 253-515, 260-781, 296-532, 302-521, 351-776, 359-621, 362-659, 363-920, 368-894, 374-908, 415-1021, 422-1021, 457-656, 474-660, 474-713, 474-1045, 485-761, 485-771, 486-721, 522-807, 538-809, 545-786, 546-994, 576-994, 582-860, 589-871, 591-1119, 613-883, 660-1000, 664-937, 720-1018, 728-1273, 782-1025, 815-1187, 843-1089, 890-1123, 890-1358, 924-1102, 965-1362, 967-1247, 1017-1296, 1043-1616, 1057-1399, 1068-1343, 1080-1366, 1088-1388, 1145-1447, 1171-1419, 1178-1295, 1180-1652, 1180-1706, 1184-2029, 1187-1351, 1192-1281, 1193-1623, 1200-1388, 1208-1342, 1221-1474, 1229-1533, 1231-1831, 1236-1802, 1237-1747, 1238-1632, 1247-1829, 1248-1433, 1256-1518, 1262-1547, 1273-1518, 1280-1533, 1284-2000, 1294-1551, 1298-1724, 1299-1878, 1301-1528, 1320-1594, 1322-1647, 1336-1596, 1336-1605, 1345-1603, 1358-1606, 1358-1897, 1363-1788, 1363-1806,
	1367-1627, 1367-1651, 1368-1640, 1368-1652, 1379-1642, 1384-1831, 1384-1953, 1387-1898, 1393-1558, 1393-2240, 1394-1689, 1395-1919, 1398-1889, 1402-1588, 1402-1830, 1406-1830, 1414-1688, 1424-2268, 1426-2189, 1431-1925, 1431-1979, 1441-1705, 1441-1734, 1441-1984, 1443-1944, 1443-2001, 1445-1683, 1453-2096, 1465-1997, 1469-2160, 1470-2052, 1481-1723, 1484-1758, 1487-1943, 1505-2198, 1510-2124, 1551-1821, 1557-2126, 1562-2095, 1567-1831, 1573-2168, 1573-2212, 1580-1911, 1581-2192, 1582-1856, 1588-1831, 1588-1859, 1590-1827, 1590-1883, 1590-2094, 1592-2053, 1598-1894, 1609-1889, 1651-2201, 1660-2021, 1664-2271, 1680-1912, 1681-1811, 1685-1998, 1685-2200, 1685-2272, 1702-2251, 1710-2215, 1711-1991, 1715-1987, 1716-2015, 1718-2345, 1733-1987, 1733-2010, 1737-2027, 1739-2253, 1741-1894, 1742-2048, 1745-2042, 1751-2015, 1760-2035, 1767-1999, 1773-2069, 1791-2215, 1793-2047, 1794-2010, 1794-2029, 1794-2157, 1796-2175, 1798-1930, 1799-2066, 1812-2077, 1812-2320, 1812-2321, 1814-1970, 1841-2359, 1842-2114, 1845-2119,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1846-2070, 1849-2119, 1857-2132, 1862-2081, 1862-2105, 1864-2198, 1865-2135, 1876-2124, 1879-2251, 1889-2251, 1889-2252, 1893-2142, 1894-2133, 1898-2109, 1898-2136, 1898-2166, 1905-2148, 1911-2192, 1934-2251, 1940-2158, 1944-2215, 1953-2166, 1964-2204, 1968-2210, 1968-2214, 1972-2261, 1976-2223, 1976-2241, 1978-2170, 1984-2225, 1988-2511, 1994-2247, 1998-2239, 1998-2242, 2006-2226, 2025-2262, 2035-2250, 2036-2583, 2045-2533, 2052-2335, 2052-2344, 2058-2253, 2059-2259, 2062-2573, 2063-2646, 2073-2366, 2074-2339, 2103-2338, 2108-2648, 2144-2568, 2151-2653, 2183-2570, 2193-2583, 2198-2393, 2209-2424, 2211-2473, 2217-2503, 2229-2656, 2248-2483, 2250-2412, 2250-2481, 2260-2476, 2261-2502, 2276-2630, 2276-2631, 2277-2619, 2279-2631, 2286-2646, 2291-2652, 2316-2537, 2329-2631, 2330-2622, 2345-2591, 2367-2691, 2370-2646, 2408-2631, 2428-2631, 2472-2643, 2472-2646, 2511-2631, 2511-2748
115/7511817CB1 1989	1-1969, 32-283, 51-314, 52-303, 157-279, 157-298, 157-300, 157-308, 157-312, 157-313, 157-314, 157-319, 157-328, 157-332, 157-346, 216-461, 342-1071, 345-592, 345-803, 345-914, 345-1052, 354-619, 355-606, 356-587, 358-694, 362-611, 365-861, 369-943, 385-640, 397-942, 402-829, 403-918, 404-565, 414-819, 415-660, 415-771, 417-658, 419-711, 434-686, 435-928, 454-751, 458-701, 476-1134, 482-749, 485-717, 485-958, 485-1029, 489-743, 498-1047, 507-1182, 511-593, 517-1060, 518-774, 528-946, 532-1088, 532-1096, 535-1123, 541-1183, 546-851, 563-1102, 564-1245, 567-941, 568-850, 585-846, 590-874, 594-1111, 601-834, 603-1213, 606-1243, 606-1267, 607-1256, 618-1209, 622-852, 627-856, 627-1301, 636-902, 651-916, 652-927, 659-813, 683-830, 705-1382, 707-954, 707-1160, 708-872, 729-1126, 729-1455, 733-933, 747-1310, 760-1033, 770-1409, 776-1159, 786-1043, 786-1377, 791-1063, 802-1061, 802-1085, 802-1435, 803-1064, 806-1371, 807-1267, 827-1375, 830-1118, 834-1097, 834-1424, 846-1450, 850-1120, 852-1093, 855-1795, 867-1110,
	873-1131, 886-1576, 891-1495, 898-1486, 900-1602, 902-1132, 906-1170, 919-1585, 922-1189, 932-1493, 940-1543, 942-1204, 942-1448, 943-1635, 949-1199, 950-1376, 952-1231, 956-1237, 957-1503, 959-1221, 960-1158, 965-1161, 983-1147, 985-1215, 986-1667, 988-1205, 990-1233, 990-1746, 993-1368, 994-1267, 996-1494, 998-1251, 1005-1441, 1007-1545, 1015-1500, 1021-1274, 1021-1306, 1022-1395, 1026-1681, 1036-1674, 1043-1637, 1050-1373, 1070-1490, 1078-1324, 1086-1226, 1086-1592, 1088-1334, 1088-1368, 1088-1371, 1096-1815, 1099-1388, 1117-1697, 1126-1727, 1131-1412, 1135-1396, 1138-1593, 1139-1846, 1152-1611, 1164-1684, 1176-1757, 1181-1455, 1186-1396, 1186-1449, 1191-1452, 1202-1459, 1208-1806, 1218-1683, 1223-1469, 1230-1681, 1233-1501, 1234-1334, 1245-1628, 1246-1525, 1250-1811, 1251-1529, 1251-1562, 1251-1573, 1273-1949, 1277-1520, 1279-1540, 1284-1816, 1292-1610, 1293-1588, 1294-1858, 1299-1822, 1309-1549, 1314-1566, 1317-1856, 1324-1444, 1326-1960, 1331-1627, 1333-1950, 1337-1612, 1340-1956, 1352-1908, 1355-1582,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1361-1589, 1363-1558, 1363-1593, 1364-1577, 1366-1637, 1369-1910, 1377-1596, 1383-1640, 1383-1980, 1389-1640, 1390-1645, 1394-1618, 1399-1660, 1400-1899, 1401-1897, 1407-1630, 1414-1652, 1414-1669, 1416-1688, 1428-1649, 1429-1835, 1443-1975, 1460-1667, 1462-1744, 1462-1747, 1462-1748, 1462-1754, 1462-1762, 1462-1772, 1462-1788, 1465-1962, 1472-1762, 1475-1890, 1476-1800, 1480-1728, 1483-1987, 1487-1959, 1489-1956, 1492-1971, 1498-1956, 1502-1956, 1504-1964, 1505-1959, 1512-1958, 1513-1970, 1515-1960, 1517-1947, 1519-1977, 1522-1989, 1526-1956, 1526-1960, 1532-1956, 1533-1874, 1534-1920, 1534-1956, 1536-1956, 1536-1957, 1538-1956, 1538-1960, 1538-1963, 1548-1956, 1551-1960, 1558-1836, 1558-1956, 1561-1963, 1562-1955, 1566-1963, 1572-1956, 1576-1958, 1579-1695, 1580-1935, 1580-1963, 1594-1957, 1597-1964, 1598-1957, 1600-1956, 1600-1958, 1602-1957, 1605-1956, 1622-1882, 1622-1914, 1623-1956, 1630-1956, 1639-1886, 1639-1970, 1641-1977, 1642-1921, 1648-1927, 1649-1958, 1673-1956, 1735-1869, 1738-1948, 1753-1964, 1794-1987, 1817-1958, 1863-1958, 1871-1960
116/7511832CB1 740	1-89, 1-740, 5-400, 7-295, 21-498, 90-339, 92-736, 94-358, 108-365, 109-395, 129-740, 131-672, 135-679, 141-457, 160-723, 167-739, 172-740, 186-670, 192-735, 195-740, 217-460, 217-476, 221-740, 224-740, 226-684, 228-668, 237-497, 243-740, 257-732, 258-728, 263-520, 263-728, 263-729, 267-740, 275-736, 277-729, 278-727, 278-731, 282-536, 283-671, 284-732, 287-725, 287-740, 288-736, 289-738, 289-740, 306-729, 310-740, 318-735, 323-736, 326-740, 327-732, 334-738, 339-700, 344-663, 360-725, 371-732, 393-651, 401-654, 402-732, 409-735, 426-732, 426-737, 429-728, 431-727, 435-728, 454-736, 465-727, 471-734, 512-722, 533-729, 640-727, 640-728
117/7512301CB1 597	1-246, 1-247, 1-597, 49-166, 49-189, 73-336, 191-597, 199-597, 200-596, 228-597, 234-439, 252-584, 318-597, 401-594, 429-594
118/7512320CB1 3113	1-423, 1-526, 1-648, 1-2697, 9-463, 9-505, 174-453, 174-460, 174-1002, 174-1017, 174-1049, 174-1113, 771-1440, 771-1581, 771-1672, 771-1708, 771-1725, 772-1434, 772-1598, 777-1591, 806-1613, 810-1532, 810-1611, 821-1630, 857-1551, 867-1640, 892-1719, 893-1437, 901-1752, 981-1637, 989-1757, 989-1758, 1007-1804, 1010-1855, 1013-1727, 1013-1791, 1019-1764, 1037-1836, 1039-1794, 1064-1819, 1072-1866, 1079-1920, 1091-1940, 1091-1957, 1091-1996, 1092-1872, 1092-1932, 1092-1935, 1092-1993, 1092-2004, 1093-1647, 1093-1936, 1102-1757, 1116-1927, 1129-1611, 1129-1612, 1129-1613, 1129-1618, 1129-1626, 1129-1718, 1129-1760, 1131-1404, 1141-1904, 1144-1933, 1149-1734, 1150-1688, 1154-1935, 1159-1928, 1169-1843, 1170-1949, 1185-1835, 1199-1957, 1223-1888, 1247-1862, 1281-1550, 1282-1984, 1295-2204, 1296-1821, 1296-1906, 1315-1898, 1328-1624, 1333-2095, 1359-1976, 1371-1924, 1382-1865, 1393-1538, 1396-1898, 1399-2293, 1401-1997, 1405-2293, 1407-2293, 1412-1933, 1416-2293, 1422-2293, 1425-2293, 1428-1621, 1428-2012,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1432-2293, 1433-2293, 1434-2293, 1463-2091, 1466-1764, 1466-2293, 1468-2293, 1469-2293, 1470-2293, 1476-2301, 1477-2266, 1478-2293, 1480-2331, 1482-2293, 1483-1952, 1485-2293, 1487-2293, 1505-2293, 1509-2293, 1517-2293, 1519-2293, 1522-2293, 1523-2061, 1524-2293, 1531-2293, 1536-2293, 1544-2293, 1550-2293, 1558-1829, 1560-2293, 1561-2170, 1561-2293, 1562-2293, 1563-2292, 1564-2293, 1568-2293, 1569-2293, 1580-2293, 1584-2293, 1587-2293, 1595-2388, 1614-1950, 1627-2293, 1630-2419, 1642-2293, 1659-2513, 1667-2427, 1724-2665, 1728-2293, 1738-2665, 1739-2665, 1746-2466, 1750-2175, 1753-2293, 1756-2308, 1757-2665, 1766-2665, 1780-2597, 1783-2386, 1785-2665, 1788-2396, 1788-2578, 1789-2386, 1790-2665, 1797-2373, 1799-2665, 1800-2660, 1802-2665, 1806-2403, 1817-2665, 1827-2407, 1827-2583, 1828-2665, 1833-2392, 1841-2585, 1842-2660, 1845-2665, 1848-2665, 1849-2665, 1851-2665, 1852-2665, 1853-2665, 1856-2441, 1856-2664, 1856-2665, 1858-2470, 1858-2665, 1864-2481, 1878-2665, 1881-2665, 1882-2649, 1892-2665, 1899-2411,
	1900-2665, 1904-2642, 1909-2518, 1912-2211, 1912-2588, 1914-2665, 1915-2665, 1920-2158, 1920-2665, 1922-2665, 1924-2665, 1932-2617, 1936-2665, 1947-2596, 1956-2665, 1963-2523, 1964-2665, 1970-2665, 1972-2665, 1973-2665, 1985-2665, 1996-2665, 2002-2665, 2014-2665, 2015-2665, 2017-2679, 2018-2661, 2021-2787, 2034-2576, 2036-2679, 2053-2293, 2055-2301, 2055-2450, 2055-2479, 2055-2574, 2055-2671, 2059-2691, 2070-2728, 2071-2643, 2071-2690, 2071-2696, 2071-2697, 2071-2698, 2071-2699, 2071-2700, 2080-2661, 2093-2686, 2102-2380, 2152-2505, 2154-2766, 2202-2665, 2219-2611, 2251-2700, 2253-2777, 2267-2456, 2267-2460, 2267-2677, 2276-2677, 2279-2547, 2289-2797, 2325-2774, 2327-2787, 2329-2642, 2562-3113
119/7512371CB1 943	1-211, 1-247, 1-943, 21-260, 21-264, 23-308, 26-333, 27-270, 27-275, 27-310, 27-327, 29-307, 29-317, 29-324, 32-285, 32-298, 33-229, 33-260, 33-284, 33-288, 33-296, 33-322, 33-554, 34-212, 34-246, 34-250, 34-295, 34-300, 35-317, 40-279, 41-239, 41-285, 43-298, 44-288, 44-316, 44-318, 44-324, 44-327, 45-287, 45-310, 48-296, 50-262, 51-304, 51-317, 51-323, 51-324, 51-333, 53-292, 53-317, 55-295, 58-317, 59-330, 60-208, 60-309, 60-313, 60-333, 62-313, 62-331, 63-329, 63-334, 64-293, 64-333, 65-178, 65-194, 65-308, 67-315, 67-317, 75-307, 75-324, 77-296, 78-304, 78-314, 78-330, 78-333, 79-213, 79-345, 80-315, 80-316, 80-328, 82-273, 82-303, 82-316, 82-317, 82-333, 83-270, 83-313, 83-317, 84-261, 84-266, 84-282, 84-310, 84-315, 84-328, 84-333, 85-316, 85-317, 86-226, 88-317, 88-333, 90-202, 90-292, 90-317, 90-333, 92-333, 93-245, 113-297, 162-297, 312-930, 326-571, 326-573, 333-361, 333-520, 333-527, 334-361, 334-722, 334-937, 335-867, 335-889, 339-802, 340-588, 340-634, 340-938, 343-722, 349-937, 361-557,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	362-597, 364-613, 368-880, 376-645, 383-720, 385-640, 385-679, 393-609, 397-920, 410-668, 411-937, 418-937, 429-943, 430-863, 434-940, 438-911, 441-833, 442-937, 446-937, 446-940, 450-937, 452-938, 453-696, 465-709, 465-748, 466-691, 466-721, 466-727, 466-937, 470-752, 475-740, 475-938, 476-937, 482-940, 483-921, 497-937, 499-937, 502-789, 506-718, 509-943, 510-943, 513-943, 516-745, 518-889, 518-893, 518-937, 519-730, 519-744, 519-770, 519-771, 519-792, 519-937, 521-941, 522-937, 522-940, 523-916, 525-767, 525-943, 528-705, 528-768, 530-937, 531-937, 537-938, 538-822, 538-871, 538-939, 540-943, 541-943, 544-943, 546-792, 550-936, 554-943, 556-831, 556-836, 557-940, 558-759, 559-794, 559-899, 559-937, 559-940, 560-778, 560-797, 560-937, 563-867, 564-839, 566-885, 567-938, 568-936, 568-940, 569-943, 575-869, 575-881, 575-937, 575-943, 576-943, 578-917, 580-937, 581-860, 581-940, 589-937, 593-937, 594-937, 607-938, 614-943, 628-925, 628-940, 640-943, 645-943, 646-940, 650-874, 651-922, 656-937, 662-943,
	665-937, 667-863, 671-874, 671-913, 672-940, 675-940, 676-939, 681-937, 695-942, 699-937, 715-938, 716-943, 728-943, 756-895, 778-937, 808-937, 838-937, 864-943
120/7512442CB1 2203	1-39, 1-723, 1-832, 1-847, 1-855, 1-884, 1-964, 2-852, 3-39, 4-708, 4-783, 59-583, 124-520, 124-527, 124-978, 124-2203, 128-783, 204-817, 211-855, 311-767, 311-822, 375-845, 422-845, 484-855, 505-837, 505-855, 539-855, 539-856, 650-873, 704-1584, 811-1581, 822-1584, 855-1359, 855-1584, 858-1345, 873-1356, 891-1367, 891-1461, 896-1358, 920-1158, 920-1338, 927-1352, 1007-1297, 1019-1274, 1021-1367, 1028-1365, 1073-1525, 1167-1358, 1167-1364, 1203-1777, 1204-1719, 1211-1737, 1217-1776, 1267-1717, 1284-1820, 1337-1777, 1370-1746, 1375-1779, 1381-1779, 1416-1779, 1432-1601, 1432-1682, 1432-1840, 1432-2004, 1436-2004, 1453-1884, 1460-1779, 1549-1860, 1557-1809, 1563-1976, 1576-2008, 1587-2013, 1674-1871, 1687-2203, 1696-1955, 1699-2002, 1702-1981, 1714-2008, 1934-2003
121/7512311CB1 1890	1-204, 1-1882, 3-247, 12-522, 24-227, 82-319, 100-348, 162-341, 171-287, 182-415, 448-704, 448-711, 449-1296, 474-714, 492-747, 493-741, 540-757, 548-858, 549-754, 561-822, 568-823, 568-1009, 583-905, 583-910, 598-823, 631-1156, 661-1124, 669-955, 711-1473, 717-1015, 729-919, 745-1003, 749-834, 750-1418, 752-1040, 754-1077, 759-1417, 769-1397, 770-1070, 778-1057, 778-1093, 778-1198, 785-1037, 790-1501, 797-1392, 803-1067, 803-1458, 804-946, 808-1383, 832-1287, 857-979, 857-1502, 868-1221, 873-1170, 874-1028, 884-1169, 884-1305, 888-1287, 907-1158, 922-1469, 946-1165, 946-1203, 959-1802, 968-1420, 986-1213, 990-1248, 991-1500, 992-1271, 995-1284, 995-1290, 998-1183, 1009-1278, 1045-1802, 1057-1293, 1057-1631, 1080-1337, 1095-1267, 1096-1371, 1096-1532, 1096-1603, 1096-1801, 1098-1806, 1129-1753, 1135-1751, 1145-1467, 1177-1868, 1180-1810, 1197-1475, 1199-1801, 1215-1498, 1233-1746, 1236-1468, 1236-1577, 1238-1531, 1238-1544, 1242-1800, 1265-1886, 1273-1564, 1273-1868, 1273-1887, 1279-1539, 1279-1557,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1279-1569, 1279-1574, 1279-1587, 1307-1580, 1307-1592, 1307-1596, 1307-1598, 1307-1602, 1307-1607, 1307-1608, 1307-1611, 1307-1613, 1307-1614, 1307-1615, 1307-1616, 1307-1619, 1307-1620, 1307-1626, 1307-1633, 1307-1800, 1307-1834, 1307-1838, 1307-1862, 1307-1885, 1307-1887, 1308-1833, 1310-1697, 1315-1882, 1316-1579, 1320-1588, 1325-1574, 1325-1654, 1328-1586, 1339-1650, 1341-1772, 1353-1641, 1369-1606, 1370-1834, 1388-1804, 1390-1815, 1392-1661, 1403-1881, 1406-1558, 1409-1888, 1421-1888, 1434-1887, 1445-1879, 1446-1888, 1451-1888, 1468-1874, 1469-1875, 1469-1878, 1471-1882, 1474-1873, 1484-1878, 1505-1888, 1516-1870, 1521-1718, 1521-1827, 1521-1849, 1539-1874, 1561-1872, 1563-1801, 1590-1807, 1590-1852, 1598-1885, 1617-1888, 1619-1876, 1622-1811, 1634-1887, 1637-1888, 1659-1890, 1670-1888, 1687-1874, 1751-1888, 1785-1888
122/7512474CB1 1035	1-236, 14-217, 24-219, 24-276, 24-278, 24-416, 33-221, 33-514, 36-926, 37-163, 37-234, 37-315, 38-225, 38-258, 38-284, 39-277, 46-551, 67-523, 69-318, 82-409, 113-267, 113-402, 133-550, 157-480, 177-345, 242-710, 595-1035, 717-796, 717-806

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
62	7510186CB1	LNOMNOF03
63	7510045CB1	NGANNOT01
66	7504544CB1	HEALDIT02
67	7504804CB1	THYRTMT01
68	7510156CB1	BRABDIR01
69	7510157CB1	BRABDIR01
70	7510993CB1	BRAIFEN05
71	7511149CB1	SININOT03
72	7511184CB1	STOMFET02
73	7511240CB1	UTRSTMR02
74	7511376CB1	PTHYNOT04
75	7501330CB1	ADRENOF04
76	7509961CB1	BRSTNOT02
77	7509963CB1	293TF2T01
78	7505089CB1	THP1NOB01
79	7510139CB1	BRAENOT04
80	7505053CB1	TLYMUNT03
81	7511116CB1	LIVRTUT04
82	7511175CB1	TLYJNOT01
83	7504660CB1	LATRTUT02
84	7504681CB1	HEAANOT01
85	7506472CB1	PROTDNV12
86	7506483CB1	THP1NOB01
87	7506525CB1	BRAITUT12
88	7506549CB1	BRAUTDR02
89	7506683CB1	KIDEUNE02
90	7510814CB1	BRAVUNT02
91	7504727CB1	THYRNOT09
92	7506958CB1	BRAINOR03
93	7505332CB1	THYRNOT03
94	7505364CB1	OSTEUNC01
95	7505455CB1	PROSNOT14
96	7505641CB1	SINITMT04
97	7511242CB1	BRAIFER05
98	7506564CB1	ISLTNOT01
99	7509076CB1	KIDCTME01
100	7506666CB1	LIVRTUN04
101	7511731CB1	LPARNOT02
102	7511735CB1	LPARNOT02
103	7511729CB1	CONUTUT01
104	7511255CB1	GBLANOT02
105	7511628CB1	LUNGNOT02
106	7511657CB1	ADRETUT05
109	7511046CB1	BRAINOT11
111	7511219CB1	DENDTNT01
112	7510933CB1	LUNGAST01
113	7511461CB1	BRSTTUT17
114	7511808CB1	SMCCNOT01
115	7511817CB1	PROSTUS23

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
116	7511832CB1	TLYMNOT08
117	7512301CB1	BRAITUT21
118	7512320CB1	UTRENOT10
119	7512371CB1	KIDNNOT19
120	7512442CB1	PLACNOB01
121	7512311CB1	PROSTUS23
122	7512474CB1	PLACNOB01

Table 6

Library	Vector	Library Description
293TF2T01	pINCY	Library was constructed using RNA isolated from a treated, transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue. The cells were treated with 5-aza-2'-deoxycytidine and transformed with adenovirus 5 DNA.
ADRENOF04	PCMV-ICIS	Library was constructed using RNA isolated from adrenal gland tissue removed from a 20-year-old Caucasian male, who died from head trauma. Serology was negative. Patient history included occasional alcohol use. Patient medications included Pepcid, Ancef, and DDAVP (antidiuretic hormone).
ADRETUT05	pINCY	Library was constructed using RNA isolated from adrenal tumor tissue removed from a 52-year-old Caucasian female during a unilateral adrenalectomy. Pathology indicated a pheochromocytoma.
BRABDIR01	pINCY	Library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Patient history included Huntington's disease, emphysema, and tobacco abuse.
BRAENOT04	pINCY	Library was constructed using RNA isolated from inferior parietal cortex tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver.
BRAIFEN05	pINCY	This normalized fetal brain tissue library was constructed from 3.26 million independent clones from a fetal brain library. Starting RNA was made from brain tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart at 23 weeks' gestation. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAIFER05	pINCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.
BRAINOR03	PBK-CMV	This random primed library was constructed using pooled cDNA from two donors. cDNA was generated using mRNA isolated from brain tissue removed from a Caucasian male fetus (donor A) who was stillborn with a hypoplastic left heart at 23 weeks' gestation and from brain tissue removed from a Caucasian male fetus (donor B), who died at 23 weeks' gestation from premature birth. Serologies were negative for both donors and family history for donor B included diabetes in the mother.
BRAINOT11	pINCY	Library was constructed using RNA isolated from brain tissue removed from the right temporal lobe of a 5-year-old Caucasian male during a hemispherectomy. Pathology indicated extensive polymicrogyria and mild to moderate gliosis (predominantly subpial and subcortical), consistent with chronic seizure disorder. Family history included a cervical neoplasm.

Table 6

Library	Vector	Library Description
BRAITUT12	pINCY	Library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 40-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated grade 4 gemistocytic astrocytoma.
BRAITUT21	pINCY	Library was constructed using RNA isolated from brain tumor tissue removed from the midline frontal lobe of a 61-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated subfrontal meningothelial meningioma with no atypia. One ethmoid and mucosal tissue sample indicated meningioma. Family history included cerebrovascular disease, senile dementia, hyperlipidemia, benign hypertension, atherosclerotic coronary artery disease, congestive heart failure, and breast cancer.
BRAUTDR02	PCDNA2.1	This random primed library was constructed using RNA isolated from pooled amygdala and entorhinal cortex tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRAVUNT02	PSPORT1	Library was constructed using pooled RNA isolated from separate populations of unstimulated astrocytes.
BRSTNOT02	PSPORT1	Library was constructed using RNA isolated from diseased breast tissue removed from a 55-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated proliferative fibrocysytic changes characterized by apocrine metaplasia, sclerosing adenosis, cyst formation, and ductal hyperplasia without atypia. Pathology for the associated tumor tissue indicated an invasive grade 4 mammary adenocarcinoma. Patient history included atrial tachycardia and a benign neoplasm. Family history included cardiovascular and cerebrovascular disease.
BRSTTUT17	pINCY	Library was constructed using RNA isolated from left breast tumor tissue removed from a 65-year-old Caucasian female during a unilateral radical mastectomy. Pathology indicated invasive and in situ grade 3, nuclear grade 2 ductal carcinoma. Patient history included hyperlipidemia and uterine leiomyoma. Family history included stomach cancer, myocardial infarction, atherosclerotic coronary artery disease, prostate cancer, benign hypertension, breast cancer, and hyperlipidemia.

Table 6

Library	Vector	Library Description
CONUTUT01	pINCY	Library was constructed using RNA isolated from sigmoid mesentery tumor tissue obtained from a 61-year-old female during a total abdominal hysterectomy and bilateral salpingo-oophorectomy with regional lymph node excision. Pathology indicated a metastatic grade 4 malignant mixed mullerian tumor present in the sigmoid mesentery at two sites.
DENDTNT01	pINCY	Library was constructed using RNA isolated from treated dendritic cells from peripheral blood.
GBLANOT02	pINCY	Library was constructed using RNA isolated from diseased gallbladder tissue removed from a 21-year-old Caucasian male during a cholecystectomy. Pathology indicated moderate chronic cholecystitis, cholelithiasis with 1 mixed stone, and acute serositis. Family history included benign hypertension, breast cancer, colon cancer, and type II diabetes.
HEAANOT01	pINCY	Library was constructed using RNA isolated from right coronary and right circumflex coronary artery tissue removed from the explanted heart of a 46-year-old Caucasian male during a heart transplantation. Patient history included myocardial infarction from total occlusion of the left anterior descending coronary artery, atherosclerotic coronary artery disease, hyperlipidemia, myocardial ischemia, dilated cardiomyopathy, left ventricular dysfunction, and tobacco abuse. Previous surgeries included cardiac catheterization. Family history included atherosclerotic coronary artery disease.
HEALDIT02	PSPORT1	Library was constructed using RNA isolated from diseased left ventricle tissue removed from a 56-year-old male during a heart transplant. Patient history included cardiovascular disease and myocardial infarction.
ISLTNOT01	pINCY	Library was constructed using RNA isolated from a pooled collection of pancreatic islet cells.
KIDCTME01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from kidney cortex tissue removed from a 65-year-old male during nephroureterectomy. Pathology indicated the margins of resection were free of involvement. Pathology for the matched tumor tissue indicated grade 3 renal cell carcinoma, clear cell type, forming a variegated multicystic mass situated within the mid-portion of the kidney. The tumor invaded deeply into but not through the renal capsule.
KIDEUNE02	pINCY	This 5' biased random primed library was constructed using RNA isolated from an untreated transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue (Invitrogen). The cells were transformed with adenovirus 5 DNA.

Table 6

Library	Vector	Library Description
KIDNNOT19	pINCY	Library was constructed using RNA isolated from kidney tissue of a 65-year-old Caucasian male during an exploratory laparotomy and nephroureterectomy. Pathology for the associated tumor tissue indicated a grade 1 renal cell carcinoma within the upper pole of the left kidney. Patient history included malignant melanoma of the abdominal skin, benign neoplasm of the colon, cerebrovascular disease, and umbilical hernia. Family history included myocardial infarction, atherosclerotic coronary artery disease, cerebrovascular disease, prostate cancer, myocardial infarction, and atherosclerotic coronary artery disease.
LATRTUT02	pINCY	Library was constructed using RNA isolated from a myxoma removed from the left atrium of a 43-year-old Caucasian male during annuloplasty. Pathology indicated atrial myxoma. Patient history included pulmonary insufficiency, acute myocardial infarction, atherosclerotic coronary artery disease, hyperlipidemia, and tobacco use. Family history included benign hypertension, acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
LIVRTUN04	pINCY	This normalized liver tumor cell line library was constructed from 1.72 million independent clones from a hepatocyte library. Starting RNA was isolated from an untreated C3A hepatocyte cell line, which is a derivative of a hepatoblastoma removed from a 15-year-old Caucasian male. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791,(except that a significantly longer (48 - hours/round) reannealing hybridization was used.
LIVRTUT04	pINCY	Library was constructed using RNA isolated from liver tumor tissue removed from a 50-year-old Caucasian male during a partial hepatectomy. Pathology indicated a grade 3-4 hepatoma, forming a mass. Patient history included benign hypertension and hepatitis. Hepatitis B core antigen and hepatitis B surface antigen was present in the patient.
LNOMNOF03	PCMV-ICIS	Library was constructed using RNA isolated from mesenteric lymph node tissue removed from a 13-year-old Caucasian male who died from intracranial bleed. Serologies were negative. Previous surgeries included tonsillectomy.
LPARNOT02	pINCY	Library was constructed using RNA isolated from tissue obtained from the left parotid (salivary) gland of a 70-year-old male with parotid cancer.
LUNGAST01	PSPORT1	Library was constructed using RNA isolated from the lung tissue of a 17-year-old Caucasian male, who died from head trauma. Patient history included asthma.
LUNGNOT02	PBLUESCRIPT	Library was constructed using RNA isolated from the lung tissue of a 47-year-old Caucasian male, who died of a subarachnoid hemorrhage.

Table 6

Library	Vector	Library Description
NGANNOT01	PSPORT1	Library was constructed using RNA isolated from tumorous neuroganglion tissue removed from a 9-year-old Caucasian male during a soft tissue excision of the chest wall. Pathology indicated a ganglioneuroma. Family history included asthma.
OSTEUNC01	pINCY	This large size-fractionated library was constructed using RNA isolated from untreated osteoblast tissue removed from the clavicle of a 40-year-old male.
PLACNOB01	PBLUESCRIPT	Library was constructed using RNA isolated from placenta.
PROSNOT14	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 60-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). Patient history included a kidney cyst and hematuria. Family history included benign hypertension, cerebrovascular disease, and arteriosclerotic coronary artery disease.
PROSTUS23	pINCY	This subtracted prostate tumor library was constructed using 10 million clones from a pooled prostate tumor library that was subjected to 2 rounds of subtractive hybridization with 10 million clones from a pooled prostate tissue library. The starting library for subtraction was constructed by pooling equal numbers of clones from 4 prostate tumor libraries using mRNA isolated from prostate tumor removed from Caucasian males at ages 58 (A), 61 (B), 66 (C), and 68 (D) during prostatectomy with lymph node excision. Pathology indicated adenocarcinoma in all donors. History included elevated PSA, induration and tobacco abuse in donor A; elevated PSA, induration, prostate hyperplasia, renal failure, osteoarthritis, renal artery stenosis, benign HTN, thrombocytopenia, hyperlipidemia, tobacco/alcohol abuse and hepatitis C (carrier) in donor B; elevated PSA, induration, and tobacco abuse in donor C; and elevated PSA, induration, hypercholesterolemia, and kidney calculus in donor D. The hybridization probe for
		subtraction was constructed by pooling equal numbers of cDNA clones from 3 prostate tissue libraries derived from prostate tissue, prostate epithelial cells, and fibroblasts from prostate stroma from 3 different donors. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR 19 (1991):1954 and Bonaldo, et al. Genome Research 6 (1996):791.

Table 6

Library	Vector	Library Description
PROTDNV12	PCR2-TOPOTA	Library was constructed using pooled cDNA from different donors. cDNA was generated using mRNA isolated from pooled small intestine tissue removed from a Caucasian male fetus (donor A) who died at 23 weeks' gestation from premature birth; from lung tissue removed from a Caucasian male fetus (donor B) who died from fetal demise; from pleura tumor tissue removed from a 55-year-old Caucasian female (donor C) during a complete pneumonectomy; from frontal/parietal brain tumor tissue removed from a 2-year-old Caucasian female (donor D) during excision of cerebral meningeal lesion; from liver tumor tissue removed from a 72-year-old Caucasian male (donor E) during partial hepatectomy; from pooled fetal brain tissue removed from a Caucasian male fetus (donor F) who was stillborn with a hypoplastic left heart at 23 weeks' gestation and from brain tissue removed from a Caucasian male fetus (donor G), who died at 23 weeks' gestation from premature birth; from pooled fetal kidney tissue removed from 59, 20-33-week-old
		male and female fetuses who died from spontaneous abortion; from pooled thymus tissue removed from 9, 18-32-year-old male and female donors who died from sudden death; and from pooled fetal liver tissue removed from 32, 18-24-week-old male and female fetuses. For donor A, serologies were negative. Family history included diabetes in the mother. For donor B, Serologies were negative. For donor C, pathology indicated grade 3 sarcoma most consistent with leiomyosarcoma, uterine primary, forming a bosellated mass replacing the right lower lobe and a portion of the middle lobe. Multiple nodules comprising the tumor show near total necrosis. Smooth muscle actin was positive. Estrogen receptor was negative and progesterone receptor was positive. The patient presented with shortness of breath. Patient history included peptic ulcer disease, normal delivery, anemia, and tobacco abuse in remission. Previous surgeries included total abdominal hysterectomy, bilateral salpingo-oophorectomy, hemorrhoidectomy, endoscopic
		excision of lung lesion, and appendectomy. Patient medications included Megace, tamoxifen, and Pepcid. Family history included multiple sclerosis in the mother; atherosclerotic coronary artery disease and type II diabetes in the father; and breast cancer in the grandparent(s). For donor D, pathology indicated neuroectodermal tumor with advanced ganglionic differentiation. The lesion was only moderately cellular but was mitotically active with a high MIB-1 labelling index. Neuronal differentiation was widespread and advanced. Multinucleate and dysplastic-appearing forms were readily seen. The glial element was less prominent. The patient presented with motor seizures. Family history included hypertension in the grandparent(s). For donor E, pathology indicated metastatic grade 2 (of 4) neuroendocrine carcinoma forming a mass. The patient presented with metastatic liver cancer. Patient history included benign hypertension, type I diabetes, prostatic hyperplasia, prostate cancer, alcohol abuse in remission, and

Table 6

Library	Vector	Library Description
PROTDNV12 - Cont'd		tobacco abuse in remission. Previous surgeries included destruction of a pancreatic lesion, closed prostatic biopsy, transurethral prostatectomy, removal of bilateral testes and total splenectomy. Patient medications included Eulexin, Hytrin, Proscar, Ecotrin, and insulin. Family history included atherosclerotic coronary artery disease and acute myocardial infarction in the mother; atherosclerotic coronary artery disease and type II diabetes in the father. For donor F and G, Serologies were negative for both donors and family history for donor G included diabetes in the mother.
PTHYNOT04	pINCY	Library was constructed using RNA isolated from parathyroid tissue removed from a 44-year-old Caucasian male during a partial parathyroidectomy. Pathology for the associated tumor tissue indicated parathyroid carcinoma (grade 1 of 4) forming a partially cystic tan mass. Both capsular and vascular invasion were present. The patient presented with unspecified parathyroid disorder. Previous surgeries included parathyroid surgery.
SININOT03	pINCY	Library was constructed using RNA isolated from ileum tissue obtained from an 8-year-old Caucasian female, who died from head trauma. Serology was positive for cytomegalovirus (CMV).
SINITMT04	pINCY	Library was constructed using RNA isolated from ileum tissue removed from a 70-year-old Caucasian female during right hemicolectomy, open liver biopsy, flexible sigmoidoscopy, colonoscopy, and permanent colostomy. Pathology for the associated tumor indicated invasive grade 2 adenocarcinoma forming an ulcerated mass, situated 2 cm distal to the ileocecal valve. Patient history included a malignant breast neoplasm, type II diabetes, hyperlipidemia, viral hepatitis, an unspecified thyroid disorder, osteoarthritis, a malignant skin neoplasm, and normal delivery. Family history included breast cancer, atherosclerotic coronary artery disease, benign hypertension, cerebrovascular disease, breast cancer, ovarian cancer, and hyperlipidemia.
SMCCNOT01	pINCY	Library was constructed using RNA isolated from smooth muscle cells removed from the coronary artery of a 3-year-old Caucasian male.
STOMFET02	pINCY	Library was constructed using RNA isolated from stomach tissue removed from a Hispanic male fetus, who died at 18 weeks' gestation.
THP1NOB01	PBLUESCRIPT	Library was constructed using RNA isolated from cultured, unstimulated THP-1 cells. THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (ref: Int. J. Cancer (1980) 26:171).
THYRNOT03	pINCY	Library was constructed using RNA isolated from thyroid tissue removed from the left thyroid of a 28-year-old Caucasian female during a complete thyroidectomy. Pathology indicated a small nodule of adenomatous hyperplasia present in the left thyroid. Pathology for the associated tumor tissue indicated dominant follicular adenoma, forming a well-encapsulated mass in the left thyroid.

Table 6

Library	Vector	Library Description
THYRNOT09	pINCY	Library was constructed using RNA isolated from diseased thyroid tissue removed from an 18-year-old Caucasian female during an unilateral thyroid lobectomy and regional lymph node excision. Pathology indicated adenomatous goiter. This was associated with a follicular adenoma of the thyroid. Family history included thyroid cancer in the father.
THYRTMT01	pINCY	Library was constructed using RNA isolated from left thyroid tissue removed from a 56-year-old Caucasian male during a unilateral thyroid lobectomy and fine needle thyroid biopsy. Pathology for the associated tumor tissue indicated medullary carcinoma invading the overlying skeletal muscle. Metastatic medullary carcinoma involved one carotid sheath lymph node (of 9), one left neck lymph node with extra nodular extension, and a central compartment node. A microscopic focus of grade 1 papillary carcinoma was identified within the right lobe of the thyroid lobe. The left thyroid vein biopsy was negative for tumor. Patient history included hyperlipidemia, headache, and atherosclerotic coronary artery disease. Family history included cerebrovascular disease, cardiovasclar disease and bone cancer.
TLYJNOT01	pINCY	Library was constructed using RNA isolated from an untreated Jurkat cell line derived from the T cells of a male. Patient history included acute T-cell leukemia.
TLYMNOT08	pINCY	The library was constructed using RNA isolated from anergicallogenic T-lymphocyte tissue removed from an adult (40-50-year old) Caucasian male.The cells were incubated for 3 days in the presence of 1 microgram/ml OKT3 mAb and 5% human serum.
TLYMUNT03	pINCY	Library was constructed using RNA isolated from untreated peripheral blood, CD8+ T-lymphocyte cell tissue removed from a 63-year-old male. The cells were isolated from buffy coat with MACS magnetic beads.
UTRENOT10	pINCY	Library was constructed using RNA isolated from pooled uterine endometiral tissue removed from three adult females during endometrial biopsy. Pathology indicated normal endometrium. All three patients were positive for Beta-3 integrin.
UTRSTMR02	PCDNA2.1	This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from endometrial tissue removed from a 32-year-old female (donor A) and using mRNA isolated from myometrium removed from a 45-year-old female (donor B) during vaginal hysterectomy and bilateral salpingo-oophorectomy. In donor A, pathology indicated the endometrium was secretory phase. The cervix showed severe dysplasia (CIN III) focally involving the squamocolumnar junction at the 1, 6 and 7 o'clock positions. Mild koilocytotic dysplasia was also identified within the cervix. In donor B, pathology for the matched tumor tissue indicated multiple (23) subserosal, intramural, and submucosal leiomyomata. Patient history included stress incontinence, extrinsic asthma without status asthmaticus and normal delivery in donor B. Family history included cerebrovascular disease, depression, and atherosclerotic coronary artery disease in donor B.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value = 1.0E-8 or less; Full Length sequences: Probability value = 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value = 1.06E-6; Assembled ESTs: fasta Identity = 95% or greater and Match length = 200 bases or greater; fastx E value = 1.0E-8 or less; Full Length sequences: fastx score = 100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value = 1.0E-3 or less

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART or TIGRFAM hits: Probability value = 1.0E-3 or less; Signal peptide hits: Score = 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score ≥ GCG specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

Table 7

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
62	7510186	3143073H1	SNP00112064	235	2564	C	C	T	noncoding	n/a	n/a	n/a	n/a
62	7510186	3143907T6	SNP00112064	123	2563	C	C	T	noncoding	n/a	n/a	n/a	n/a
62	7510186	4224128F6	SNP00112063	360	2260	T	T	G	noncoding	0.45	0.35	0.77	0.45
62	7510186	4710260H1	SNP00112063	200	2261	G	T	G	noncoding	0.45	0.35	0.77	0.45
63	7510045	2677038H1	SNP00076253	127	482	G	G	A	L158	n/d	n/d	n/d	n/d
63	7510045	5184008H1	SNP00076253	153	481	G	G	A	R158	n/d	n/d	n/d	n/d
65	7510450	1440148H1	SNP00112792	145	247	C	C	T	noncoding	0.12	n/a	n/a	n/a
65	7510450	4724606H2	SNP00112792	214	245	C	C	T	noncoding	0.12	n/a	n/a	n/a
65	7510450	4799875H1	SNP00112792	242	246	C	C	T	noncoding	0.12	n/a	n/a	n/a
65	7510450	6401991H1	SNP00112792	27	237	C	C	T	noncoding	0.12	n/a	n/a	n/a
65	7510450	7926072H1	SNP00063887	458	317	A	A	G	noncoding	n/a	n/a	n/a	n/a
66	7504544	000483H1	SNP00061084	281	577	C	C	T	noncoding	n/a	n/a	n/a	n/a
66	7504544	1805764H1	SNP00013151	8	12	C	C	T	noncoding	0.28	n/a	n/a	n/a
66	7504544	2009335H1	SNP00128049	86	301	T	T	C	noncoding	n/a	n/a	n/a	n/a
66	7504544	2829760H1	SNP00061084	267	576	C	C	T	noncoding	n/a	n/a	n/a	n/a
66	7504544	3054701H1	SNP00128049	68	298	C	T	C	noncoding	n/a	n/a	n/a	n/a
66	7504544	3270496H1	SNP00013151	5	11	C	C	T	noncoding	0.28	n/a	n/a	n/a
66	7504544	3442243H1	SNP00061084	187	575	C	C	T	noncoding	n/a	n/a	n/a	n/a
66	7504544	4074015H1	SNP00128049	195	299	T	T	C	noncoding	n/a	n/a	n/a	n/a
66	7504544	4091844H1	SNP00013151	9	9	C	C	T	noncoding	0.28	n/a	n/a	n/a
66	7504544	4220878H1	SNP00128049	141	296	T	T	C	noncoding	n/a	n/a	n/a	n/a
66	7504544	6724862H1	SNP00100143	40	125	C	A	C	P15	n/a	n/a	n/a	n/a
67	7504804	4677757H1	SNP00100736	20	435	T	T	G	L134	n/a	n/a	n/a	n/a
67	7504804	6482171H1	SNP00100736	268	436	T	T	G	L134	n/a	n/a	n/a	n/a
69	7510157	7114262H1	SNP00116640	510	574	A	A	C	noncoding	n/d	n/d	n/d	n/d
70	7510993	000901H1	SNP00124903	196	1189	T	T	C	noncoding	n/a	n/a	n/a	n/a
70	7510993	000901H1	SNP00128318	271	1264	C	C	T	noncoding	n/a	n/a	n/a	n/a
70	7510993	030787H1	SNP00130018	34	519	C	C	T	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
70	7510993	037869H1	SNP00144513	151	990	C	C	T	noncoding	n/a	n/a	n/a	n/a
70	7510993	1473395H1	SNP00129079	163	1104	C	C	G	noncoding	n/d	n/a	n/a	n/a
70	7510993	1574675H1	SNP00128364	37	1521	C	C	A	noncoding	n/a	n/a	n/a	n/a
70	7510993	1851245H1	SNP00066761	195	302	C	C	G	D13	n/a	n/a	n/a	n/a
70	7510993	2109514H1	SNP00147823	136	679	C	C	T	noncoding	n/a	n/a	n/a	n/a
70	7510993	2131852T6	SNP00124903	198	1204	T	T	C	noncoding	n/a	n/a	n/a	n/a
70	7510993	2131852T6	SNP00128318	123	1279	C	C	T	noncoding	n/a	n/a	n/a	n/a
70	7510993	2368939T6	SNP00124903	222	1242	T	T	C	noncoding	n/a	n/a	n/a	n/a
70	7510993	2368939T6	SNP00128318	147	1317	C	C	T	noncoding	n/a	n/a	n/a	n/a
70	7510993	2447365T6	SNP00124903	274	1190	T	T	C	noncoding	n/a	n/a	n/a	n/a
70	7510993	2447365T6	SNP00128318	199	1265	C	C	T	noncoding	n/a	n/a	n/a	n/a
70	7510993	2453815H1	SNP00066762	198	765	C	C	T	noncoding	n/a	n/a	n/a	n/a
70	7510993	2747003H1	SNP00005155	123	474	C	C	T	noncoding	n/a	n/a	n/a	n/a
70	7510993	2747003H1	SNP00027184	110	461	C	C	T	noncoding	n/a	n/a	n/a	n/a
70	7510993	3747901T6	SNP00124903	273	1191	T	T	C	noncoding	n/a	n/a	n/a	n/a
70	7510993	3747901T6	SNP00128318	198	1266	C	C	T	noncoding	n/a	n/a	n/a	n/a
70	7510993	3890417H1	SNP00001540	188	777	T	G	T	noncoding	n/a	n/a	n/a	n/a
70	7510993	4549051T1	SNP00124903	255	1209	T	T	C	noncoding	n/a	n/a	n/a	n/a
70	7510993	4549051T1	SNP00128318	180	1284	C	C	T	noncoding	n/a	n/a	n/a	n/a
70	7510993	7676454H1	SNP00144513	275	989	C	C	T	noncoding	n/a	n/a	n/a	n/a
71	7511149	1266935F1	SNP00059946	264	1346	G	G	A	noncoding	n/a	n/a	n/a	n/a
71	7511149	135151H1	SNP00032857	72	376	C	T	C	noncoding	0.9	n/a	n/a	n/a
71	7511149	1923069H1	SNP00032860	43	1678	G	G	A	noncoding	n/a	n/a	n/a	n/a
71	7511149	2739264F6	SNP00071112	58	70	C	C	T	T18	n/a	n/a	n/a	n/a
71	7511149	2739264T6	SNP00071112	58	1819	G	G	A	noncoding	n/a	n/a	n/a	n/a
71	7511149	2744648T6	SNP00032860	245	1679	A	G	A	noncoding	n/a	n/a	n/a	n/a
71	7511149	2805590F6	SNP00032861	315	1934	C	C	T	noncoding	n/a	n/a	n/a	n/a
71	7511149	2911923H1	SNP00151372	156	238	T	T	C	V74	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
71	7511149	5092648H1	SNP00059945	177	204	T	T	C	V63	n/a	n/a	n/a	n/a
71	7511149	7441862H1	SNP00059945	152	205	T	T	C	V63	n/a	n/a	n/a	n/a
72	7511184	1580527H1	SNP00016260	66	1448	A	A	G	noncoding	n/d	n/a	n/a	n/a
72	7511184	1736429T6	SNP00099846	138	3062	G	G	T	noncoding	n/d	n/d	n/d	n/d
72	7511184	4180123H1	SNP00014085	203	3108	G	G	T	noncoding	n/a	n/a	n/a	n/a
72	7511184	4220710H1	SNP00099846	94	3081	G	G	T	noncoding	n/d	n/d	n/d	n/d
73	7511240	2530062H1	SNP00015584	69	74	C	T	C	noncoding	0.56	0.37	0.2	0.3
73	7511240	2530062H1	SNP00149856	194	199	T	T	C	F9	n/a	n/a	n/a	n/a
73	7511240	6167536T8	SNP00074803	132	1228	C	C	A	noncoding	n/a	n/a	n/a	n/a
74	7511376	1731463H1	SNP00028004	33	371	G	G	A	noncoding	n/d	n/d	n/d	n/d
74	7511376	1731463H1	SNP00062324	144	482	C	A	C	noncoding	n/a	n/a	n/a	n/a
74	7511376	1739102H1	SNP00098644	170	758	G	G	C	noncoding	n/a	n/a	n/a	n/a
74	7511376	3337377H1	SNP00098643	9	635	C	C	T	noncoding	n/a	n/a	n/a	n/a
75	7501330	1321803H1	SNP00098943	38	298	A	A	G	noncoding	n/a	n/a	n/a	n/a
75	7501330	3106005H1	SNP00098943	85	295	A	A	G	noncoding	n/a	n/a	n/a	n/a
75	7501330	3593610H1	SNP00098943	320	299	A	A	G	noncoding	n/a	n/a	n/a	n/a
75	7501330	7614239H1	SNP00040710	553	1346	G	T	G	noncoding	n/a	n/a	n/a	n/a
76	7509961	1338467H1	SNP00141303	116	770	C	C	A	noncoding	n/d	n/a	n/a	n/a
76	7509961	1345892H1	SNP00131420	115	982	T	T	G	noncoding	n/a	n/a	n/a	n/a
76	7509961	2049849H1	SNP00073984	64	467	G	G	C	E118	n/d	n/d	n/d	n/d
76	7509961	2415246H1	SNP00032099	149	535	C	C	G	P141	0.83	0.47	0.25	0.77
76	7509961	2455286H1	SNP00003178	159	242	C	C	G	L43	0.69	n/a	n/a	n/a
76	7509961	3087345H2	SNP00141303	17	769	C	C	A	noncoding	n/d	n/a	n/a	n/a
76	7509961	3323775H1	SNP00131420	254	981	T	T	G	noncoding	n/a	n/a	n/a	n/a
76	7509961	3559207H1	SNP00032099	92	534	C	C	G	P141	0.83	0.47	0.25	0.77
76	7509961	3777028H1	SNP00141303	122	768	C	C	A	noncoding	n/d	n/a	n/a	n/a
76	7509961	3918128H1	SNP00131420	143	980	T	T	G	noncoding	n/a	n/a	n/a	n/a
76	7509961	3937422H1	SNP00032099	122	533	C	C	G	F140	0.83	0.47	0.25	0.77

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
76	7509961	4158435H1	SNP00003178	159	240	C	C	G	L43	0.69	n/a	n/a	n/a
76	7509961	4321325H1	SNP00003178	6	241	C	C	G	P43	0.69	n/a	n/a	n/a
76	7509961	4570283H1	SNP00141303	57	767	C	C	A	noncoding	n/d	n/a	n/a	n/a
76	7509961	4598766H1	SNP00141303	105	766	C	C	A	noncoding	n/d	n/a	n/a	n/a
76	7509961	4786401H1	SNP00003178	156	238	C	C	G	P42	0.69	n/a	n/a	n/a
76	7509961	4828290H1	SNP00131420	96	978	T	T	G	noncoding	n/a	n/a	n/a	n/a
76	7509961	6000713H1	SNP00131420	157	976	T	T	G	noncoding	n/a	n/a	n/a	n/a
77	7509963	1284735H1	SNP00141303	161	573	C	C	A	P154	n/d	n/a	n/a	n/a
77	7509963	1338467H1	SNP00141303	116	574	C	C	A	T154	n/d	n/a	n/a	n/a
77	7509963	1345892H1	SNP00131420	115	786	T	T	G	W225	n/a	n/a	n/a	n/a
77	7509963	2455286H1	SNP00003178	159	242	C	C	G	L43	0.69	n/a	n/a	n/a
77	7509963	3323775H1	SNP00131420	254	785	T	T	G	S224	n/a	n/a	n/a	n/a
77	7509963	3393066H1	SNP00141303	220	572	C	C	A	P153	n/d	n/a	n/a	n/a
77	7509963	4158435H1	SNP00003178	159	240	C	C	G	L43	0.69	n/a	n/a	n/a
77	7509963	4321325H1	SNP00003178	6	241	C	C	G	P43	0.69	n/a	n/a	n/a
77	7509963	4570283H1	SNP00141303	57	571	C	C	A	P153	n/d	n/a	n/a	n/a
77	7509963	4598766H1	SNP00141303	105	570	C	C	A	P153	n/d	n/a	n/a	n/a
77	7509963	4786401H1	SNP00003178	156	238	C	C	G	P42	0.69	n/a	n/a	n/a
77	7509963	5295182H1	SNP00131420	7	784	T	T	G	M224	n/a	n/a	n/a	n/a
77	7509963	6000713H1	SNP00131420	157	780	T	T	G	L223	n/a	n/a	n/a	n/a
77	7509963	621891H1	SNP00141303	169	575	C	C	A	T154	n/d	n/a	n/a	n/a
78	7505089	030692H1	SNP00124454	96	866	C	C	T	H217	n/a	n/a	n/a	n/a
78	7505089	034105H1	SNP00067146	117	443	C	C	T	P76	n/a	n/a	n/a	n/a
78	7505089	1286313H1	SNP00092908	92	1175	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	7505089	154958H1	SNP00018343	135	611	C	C	T	R132	n/a	n/a	n/a	n/a
78	7505089	2399811H1	SNP00124454	11	864	C	C	T	T216	n/a	n/a	n/a	n/a
78	7505089	259782H1	SNP00067146	134	444	C	C	T	P76	n/a	n/a	n/a	n/a
78	7505089	3109003H1	SNP00124454	138	865	C	C	T	N216	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
78	7505089	3158872H1	SNP00124454	162	862	C	C	T	A215	n/a	n/a	n/a	n/a
78	7505089	4108913H1	SNP00124454	176	858	C	C	T	P214	n/a	n/a	n/a	n/a
78	7505089	4195074H1	SNP00067146	181	442	C	C	T	D75	n/a	n/a	n/a	n/a
78	7505089	4211203H1	SNP00124454	182	863	C	C	T	Q216	n/a	n/a	n/a	n/a
78	7505089	4531755H1	SNP00092908	84	1172	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	7505089	4566729H1	SNP00067146	228	441	C	C	T	A75	n/a	n/a	n/a	n/a
78	7505089	4703146H1	SNP00067146	94	437	T	C	T	Y74	n/a	n/a	n/a	n/a
78	7505089	4793334H1	SNP00067146	189	445	C	C	T	P76	n/a	n/a	n/a	n/a
78	7505089	4858472H1	SNP00067146	114	438	C	C	T	S74	n/a	n/a	n/a	n/a
78	7505089	5699517H1	SNP00124454	195	860	C	C	T	P215	n/a	n/a	n/a	n/a
78	7505089	6447625H1	SNP00018343	449	606	C	C	T	A130	n/a	n/a	n/a	n/a
78	7505089	6737674H1	SNP00124454	149	817	C	C	T	T200	n/a	n/a	n/a	n/a
78	7505089	879871H1	SNP00092908	49	1173	G	G	A	noncoding	n/a	n/a	n/a	n/a
79	7510139	1229414H1	SNP00095889	204	1112	G	G	A	noncoding	n/a	n/a	n/a	n/a
79	7510139	1229602H1	SNP00032689	125	1271	C	C	G	noncoding	n/a	n/a	n/a	n/a
79	7510139	1229602H1	SNP00032690	176	1322	G	G	C	noncoding	n/a	n/a	n/a	n/a
79	7510139	1396816H1	SNP00060055	55	1215	C	C	T	noncoding	n/d	n/a	n/a	n/a
79	7510139	1396816H1	SNP00095889	159	1111	G	G	A	noncoding	n/a	n/a	n/a	n/a
79	7510139	2152396H1	SNP00007323	206	526	C	T	C	P160	n/a	n/a	n/a	n/a
79	7510139	2152396H1	SNP00032688	93	413	C	C	T	H122	n/d	n/d	n/d	n/d
79	7510139	2420226H1	SNP00060055	162	1216	C	C	T	noncoding	n/d	n/a	n/a	n/a
79	7510139	2421485H1	SNP00007324	122	1490	C	A	C	noncoding	0.48	0.25	0.33	0.44
79	7510139	2535511H1	SNP00007323	140	528	C	T	C	L161	n/a	n/a	n/a	n/a
79	7510139	3866712H1	SNP00007323	251	527	T	T	C	L160	n/a	n/a	n/a	n/a
79	7510139	4176945H1	SNP00032688	146	417	C	C	T	R124	n/d	n/d	n/d	n/d
79	7510139	4178634H1	SNP00007323	159	525	C	T	C	L160	n/a	n/a	n/a	n/a
79	7510139	4178634H1	SNP00032688	46	412	T	C	T	L122	n/d	n/d	n/d	n/d
79	7510139	4769683H1	SNP00007324	96	1489	A	A	C	noncoding	0.48	0.25	0.33	0.44

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
79	7510139	4836961H1	SNP00007324	249	1482	A	A	C	noncoding	0.48	0.25	0.33	0.44
79	7510139	4836961H1	SNP00032689	30	1263	G	C	G	noncoding	n/a	n/a	n/a	n/a
79	7510139	4836961H1	SNP00032690	82	1314	G	G	C	noncoding	n/a	n/a	n/a	n/a
79	7510139	5204625H1	SNP00007324	200	1487	A	A	C	noncoding	0.48	0.25	0.33	0.44
79	7510139	5204625H1	SNP00032690	32	1320	G	G	C	noncoding	n/a	n/a	n/a	n/a
79	7510139	5781423H1	SNP00060055	189	1213	C	C	T	noncoding	n/d	n/a	n/a	n/a
79	7510139	5840863H2	SNP00032688	192	341	C	C	T	A98	n/d	n/d	n/d	n/d
79	7510139	5996483H1	SNP00032689	393	1171	C	C	G	noncoding	n/a	n/a	n/a	n/a
79	7510139	5996483H1	SNP00095889	234	1002	G	G	A	noncoding	n/a	n/a	n/a	n/a
79	7510139	5997065H1	SNP00032689	395	1173	C	C	G	noncoding	n/a	n/a	n/a	n/a
79	7510139	5997065H1	SNP00032690	446	1223	G	G	C	noncoding	n/a	n/a	n/a	n/a
79	7510139	5997065H1	SNP00060055	340	1111	C	C	T	noncoding	n/d	n/a	n/a	n/a
79	7510139	5997065H1	SNP00095889	236	1003	G	G	A	noncoding	n/a	n/a	n/a	n/a
79	7510139	6200424H1	SNP00095889	344	1014	G	G	A	noncoding	n/a	n/a	n/a	n/a
79	7510139	6908004J1	SNP00032688	429	386	C	C	T	N113	n/d	n/d	n/d	n/d
79	7510139	8621125J1	SNP00060055	546	1114	C	C	T	noncoding	n/d	n/a	n/a	n/a
79	7510139	8621125J1	SNP00095889	443	1007	G	G	A	noncoding	n/a	n/a	n/a	n/a
80	7505053	019957H1	SNP00009698	196	407	G	G	T	noncoding	n/d	n/a	n/a	n/a
80	7505053	1461447H1	SNP00112710	118	993	T	T	G	noncoding	n/d	n/d	1	n/d
80	7505053	1519266H1	SNP00037964	94	793	G	G	A	noncoding	n/d	n/a	n/a	n/a
80	7505053	1994332R6	SNP00037964	138	795	G	G	A	noncoding	n/d	n/a	n/a	n/a
81	7511116	1511432H1	SNP00121053	156	2131	C	C	T	noncoding	n/a	n/a	n/a	n/a
81	7511116	1725801H1	SNP00115974	155	432	A	A	G	S118	n/a	n/a	n/a	n/a
81	7511116	1921979H1	SNP00121052	46	1855	A	A	G	noncoding	n/a	n/a	n/a	n/a
81	7511116	2216279H1	SNP00135641	147	1783	C	C	T	noncoding	n/a	n/a	n/a	n/a
81	7511116	2516129F6	SNP00016548	98	98	G	G	A	P6	n/a	n/a	n/a	n/a
81	7511116	2516129F6	SNP00016549	197	197	T	T	C	P39	n/d	n/a	n/a	n/a
81	7511116	2516129H1	SNP00016549	198	216	T	T	C	C46	n/d	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
81	7511116	2853280H1	SNP00016548	129	117	G	G	A	V13	n/a	n/a	n/a	n/a
81	7511116	3041462F6	SNP00121051	83	1584	T	T	C	noncoding	n/a	n/a	n/a	n/a
81	7511116	585913IH1	SNP00121050	80	1076	A	A	G	noncoding	n/a	n/a	n/a	n/a
81	7511116	6269140H1	SNP00115975	442	701	C	C	T	S207	n/a	n/a	n/a	n/a
81	7511116	6715827H1	SNP00147432	162	1973	G	G	A	noncoding	n/a	n/a	n/a	n/a
81	7511116	763838SJ1	SNP00016548	181	119	G	G	A	V13	n/a	n/a	n/a	n/a
81	7511116	763838SJ1	SNP00016549	82	218	T	T	C	C46	n/d	n/a	n/a	n/a
82	7511175	1650336H1	SNP00033743	17	36	G	G	A	noncoding	n/a	n/a	n/a	n/a
82	7511175	1650336H1	SNP00033744	197	216	C	C	A	L40	n/d	n/d	n/d	n/d
82	7511175	2478988H1	SNP00007908	68	671	T	T	C	N191	n/a	n/a	n/a	n/a
82	7511175	3182329F6	SNP00033744	216	217	C	C	A	P40	n/d	n/d	n/d	n/d
82	7511175	3182329T6	SNP00007908	440	672	T	T	C	S192	n/a	n/a	n/a	n/a
83	7504660	007852H1	SNP00021613	228	718	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7504660	1275748H1	SNP00139585	198	253	C	C	T	A22	n/a	n/a	n/a	n/a
83	7504660	1313404H1	SNP00148065	179	279	G	G	A	V31	n/a	n/a	n/a	n/a
83	7504660	1350607H1	SNP00139744	125	229	C	C	T	A14	n/a	n/a	n/a	n/a
83	7504660	175288 1H1	SNP00021613	237	717	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7504660	228368H1	SNP00112829	152	213	A	A	G	M9	n/a	n/a	n/a	n/a
83	7504660	2317290H1	SNP00139585	161	252	C	C	T	P22	n/a	n/a	n/a	n/a
83	7504660	2972215H2	SNP00112829	65	211	A	A	G	E8	n/a	n/a	n/a	n/a
83	7504660	3713293H1	SNP00112829	35	210	A	A	G	T8	n/a	n/a	n/a	n/a
83	7504660	3963739H1	SNP00021613	259	713	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7504660	4243008H1	SNP00139585	343	251	C	C	T	V21	n/a	n/a	n/a	n/a
83	7504660	4447067H1	SNP00139585	102	244	C	C	T	T19	n/a	n/a	n/a	n/a
83	7504660	4572568H1	SNP00021613	225	716	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7504660	4575893H1	SNP00139744	168	228	C	C	T	P14	n/a	n/a	n/a	n/a
83	7504660	4593766H1	SNP00139585	119	250	C	C	T	A21	n/a	n/a	n/a	n/a
83	7504660	4643262H1	SNP00021613	201	714	C	C	T	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
83	7504660	4648505H1	SNP00148065	191	278	G	G	A	E30	n/a	n/a	n/a	n/a
83	7504660	4650084H1	SNP00139744	146	227	C	C	T	F13	n/a	n/a	n/a	n/a
83	7504660	4669396H1	SNP00139585	207	247	C	C	T	T20	n/a	n/a	n/a	n/a
83	7504660	5339825H1	SNP00112829	88	175	A	A	G	noncoding	n/a	n/a	n/a	n/a
83	7504660	5941078H1	SNP00139585	249	249	C	C	T	L21	n/a	n/a	n/a	n/a
83	7504660	6213274H1	SNP00139585	155	248	C	C	T	T20	n/a	n/a	n/a	n/a
83	7504660	6834866H1	SNP00021612	517	708	T	T	C	noncoding	n/a	n/a	n/a	n/a
84	7504681	1711158H1	SNP00067648	180	2917	C	C	T	noncoding	n/a	n/a	n/a	n/a
84	7504681	2698995H1	SNP00067648	66	2908	C	C	T	noncoding	n/a	n/a	n/a	n/a
84	7504681	4357569H1	SNP00067648	199	2915	C	C	T	noncoding	n/a	n/a	n/a	n/a
84	7504681	4654604H1	SNP00067648	188	2916	C	C	T	noncoding	n/a	n/a	n/a	n/a
85	7506472	1609713H1	SNP00023780	27	577	C	T	C	noncoding	0.85	n/a	n/a	n/a
85	7506472	2062012H1	SNP00066406	3	876	C	C	T	noncoding	n/a	n/a	n/a	n/a
85	7506472	2258417H1	SNP00023780	65	580	T	T	C	noncoding	0.85	n/a	n/a	n/a
85	7506472	2461995H1	SNP00066406	6	879	C	C	T	noncoding	n/a	n/a	n/a	n/a
85	7506472	3281958H1	SNP00066405	132	748	T	T	C	noncoding	n/a	n/a	n/a	n/a
85	7506472	6497747H1	SNP00066405	347	751	T	T	C	noncoding	n/a	n/a	n/a	n/a
85	7506472	6822056H1	SNP00098096	193	125	A	A	G	L29	n/a	n/a	n/a	n/a
85	7506472	808636H1	SNP00066406	78	873	C	C	T	noncoding	n/a	n/a	n/a	n/a
85	7506472	914139H1	SNP00023780	81	578	T	T	C	noncoding	0.85	n/a	n/a	n/a
86	7506483	030692H1	SNP00124454	96	700	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7506483	1286313H1	SNP00092908	92	1009	G	G	A	noncoding	n/a	n/a	n/a	n/a
86	7506483	154958H1	SNP00018343	135	445	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7506483	2399811H1	SNP00124454	11	698	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7506483	3109003H1	SNP00124454	138	699	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7506483	3158872H1	SNP00124454	162	696	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7506483	4211203H1	SNP00124454	182	697	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7506483	4531755H1	SNP00092908	84	1006	G	G	A	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
86	7506483	5699517H1	SNP00124454	195	694	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7506483	6447625H1	SNP00018343	449	440	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7506483	6737674H1	SNP00124454	149	651	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7506483	6929007H1	SNP00018343	366	444	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7506483	8630315J1	SNP00018343	586	458	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7506483	879871H1	SNP00092908	49	1007	G	G	A	noncoding	n/a	n/a	n/a	n/a
87	7506525	2755946H1	SNP00051200	169	1583	C	C	T	noncoding	n/a	n/a	n/a	n/a
87	7506525	2755946H1	SNP00051201	178	1574	C	C	T	noncoding	0.98	n/a	n/a	n/a
87	7506525	5377716H1	SNP00051200	38	1582	C	C	T	noncoding	n/a	n/a	n/a	n/a
87	7506525	5377716H1	SNP00051201	47	1573	C	C	T	noncoding	0.98	n/a	n/a	n/a
87	7506525	6553201H1	SNP00105530	268	1772	A	A	G	noncoding	0.78	0.93	0.91	0.74
88	7506549	7581960H1	SNP00014698	31	3226	C	C	T	noncoding	n/a	n/a	n/a	n/a
90	7510814	2432516H1	SNP00026362	61	24	A	A	G	noncoding	n/a	n/a	n/a	n/a
90	7510814	2432609H1	SNP00056302	150	1942	A	A	G	noncoding	n/d	n/a	n/a	n/a
90	7510814	2655227H1	SNP00131351	257	2879	C	T	C	noncoding	n/a	n/a	n/a	n/a
90	7510814	3223832H1	SNP00131351	221	2877	C	T	C	noncoding	n/a	n/a	n/a	n/a
90	7510814	3223832H1	SNP00131352	271	2927	C	C	A	noncoding	n/a	n/a	n/a	n/a
90	7510814	7169295H1	SNP00131352	437	2929	C	C	A	noncoding	n/a	n/a	n/a	n/a
91	7504727	1712157T6	SNP00059144	259	1497	C	C	T	noncoding	n/a	n/a	n/a	n/a
91	7504727	1861148F6	SNP00059144	54	1314	C	C	T	noncoding	n/a	n/a	n/a	n/a
91	7504727	2040761H1	SNP00007033	159	1463	C	C	T	noncoding	n/a	n/a	n/a	n/a
91	7504727	2040761H1	SNP00007034	219	1523	C	C	A	noncoding	n/a	n/a	n/a	n/a
91	7504727	411759T6	SNP00007033	317	1465	T	C	T	noncoding	n/a	n/a	n/a	n/a
91	7504727	411759T6	SNP00007034	257	1525	C	C	A	noncoding	n/a	n/a	n/a	n/a
91	7504727	411759T6	SNP00007035	131	1651	C	T	C	noncoding	n/a	n/a	n/a	n/a
91	7504727	4342951T8	SNP00007035	77	1649	T	T	C	noncoding	n/a	n/a	n/a	n/a
92	7506958	7233773H1	SNP00112954	137	617	A	G	A	noncoding	n/a	n/a	n/a	n/a
94	7505364	1235086H1	SNP00130459	197	1409	C	C	T	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
94	7505364	1403510H1	SNP00139261	23	504	G	G	C	noncoding	n/a	n/a	n/a	n/a
94	7505364	2331002H1	SNP00130459	186	1402	C	C	T	noncoding	n/a	n/a	n/a	n/a
94	7505364	2512454H1	SNP00130459	95	1408	C	C	T	noncoding	n/a	n/a	n/a	n/a
94	7505364	3869239H1	SNP00130459	99	1407	C	C	T	noncoding	n/a	n/a	n/a	n/a
94	7505364	4638567H1	SNP00139261	40	502	C	G	C	noncoding	n/a	n/a	n/a	n/a
94	7505364	5855320H1	SNP00130459	236	1404	C	C	T	noncoding	n/a	n/a	n/a	n/a
94	7505364	5954202H1	SNP00130459	173	1406	C	C	T	noncoding	n/a	n/a	n/a	n/a
94	7505364	6147443H1	SNP00124524	365	406	T	T	G	T19	n/a	n/a	n/a	n/a
95	7505455	1725135H1	SNP00112800	126	534	T	T	C	noncoding	n/a	n/a	n/a	n/a
95	7505455	2863738H1	SNP00112800	200	530	T	T	C	noncoding	n/a	n/a	n/a	n/a
95	7505455	2866674H1	SNP00146126	90	983	G	G	A	noncoding	n/a	n/a	n/a	n/a
95	7505455	2866674H1	SNP00146127	144	1037	T	T	C	noncoding	n/a	n/a	n/a	n/a
95	7505455	4024018H1	SNP00019135	50	698	G	G	A	noncoding	n/a	n/a	n/a	n/a
95	7505455	4121035H1	SNP00019135	209	699	A	G	A	noncoding	n/a	n/a	n/a	n/a
95	7505455	4121035H1	SNP00112800	42	532	T	T	C	noncoding	n/a	n/a	n/a	n/a
95	7505455	5261894H1	SNP00146126	141	985	G	G	A	noncoding	n/a	n/a	n/a	n/a
95	7505455	6249006H1	SNP00019135	187	701	G	G	A	noncoding	n/a	n/a	n/a	n/a
95	7505455	6253004H1	SNP00019135	175	690	G	G	A	noncoding	n/a	n/a	n/a	n/a
95	7505455	6253004H1	SNP00112800	8	523	T	T	C	noncoding	n/a	n/a	n/a	n/a
95	7505455	6444245H1	SNP00146125	246	285	T	C	T	noncoding	n/a	n/a	n/a	n/a
95	7505455	6894208H1	SNP00146127	73	1039	T	T	C	noncoding	n/a	n/a	n/a	n/a
96	7505641	1284528H1	SNP00010435	197	228	C	C	T	T59	n/a	n/a	n/a	n/a
96	7505641	1284528H1	SNP00040564	84	115	G	G	A	W21	n/d	n/a	n/a	n/a
96	7505641	2058585H1	SNP00040565	78	775	C	C	G	noncoding	n/a	n/a	n/a	n/a
96	7505641	2634754H1	SNP00058891	115	514	C	C	T	noncoding	n/a	n/a	n/a	n/a
96	7505641	2680212H1	SNP00010435	71	227	C	C	T	P59	n/a	n/a	n/a	n/a
96	7505641	2789884H2	SNP00010435	195	226	C	C	T	V58	n/a	n/a	n/a	n/a
96	7505641	2789884H2	SNP00040564	82	114	G	G	A	W21	n/d	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
96	7505641	4558419H1	SNP00058891	156	511	C	C	T	noncoding	n/a	n/a	n/a	n/a
96	7505641	4611562H1	SNP00040565	137	772	C	C	G	noncoding	n/a	n/a	n/a	n/a
96	7505641	4871927H1	SNP00040565	141	773	C	C	G	noncoding	n/a	n/a	n/a	n/a
96	7505641	5943093H1	SNP00040564	120	113	G	G	A	G21	n/d	n/a	n/a	n/a
96	7505641	6117637H1	SNP00058891	179	510	C	C	T	noncoding	n/a	n/a	n/a	n/a
98	7506564	1530922H1	SNP00020594	116	960	C	C	T	noncoding	n/a	n/a	n/a	n/a
98	7506564	2070458H1	SNP00111887	46	710	T	T	C	noncoding	n/a	n/a	n/a	n/a
98	7506564	2076304H1	SNP00020593	39	735	T	T	C	noncoding	n/a	n/a	n/a	n/a
99	7509076	171471H1	SNP00062685	208	1212	C	G	C	noncoding	n/a	n/a	n/a	n/a
99	7509076	1723428H1	SNP00108625	88	1027	G	G	T	noncoding	n/a	n/a	n/a	n/a
99	7509076	2212164H1	SNP00069043	221	290	C	C	T	P60	n/d	n/d	n/d	n/d
99	7509076	2500726T6	SNP00062685	116	1214	G	G	C	noncoding	n/a	n/a	n/a	n/a
99	7509076	3984515H1	SNP00037991	164	369	A	A	G	K86	n/a	n/a	n/a	n/a
99	7509076	608210R6	SNP00062685	210	1213	G	G	C	noncoding	n/a	n/a	n/a	n/a
99	7509076	608210T6	SNP00108625	323	1023	G	G	T	noncoding	n/a	n/a	n/a	n/a
100	7506666	4172221H1	SNP00095724	117	927	G	G	A	E286	n/a	n/a	n/a	n/a
101	7511731	1322927H1	SNP00036077	83	415	T	T	C	S130	n/a	n/a	n/a	n/a
101	7511731	1322983H1	SNP00066658	167	602	C	C	A	P192	n/a	n/a	n/a	n/a
101	7511731	1325778H1	SNP00036078	63	540	C	C	T	P171	n/a	n/a	n/a	n/a
102	7511735	1322927H1	SNP00036077	83	471	T	T	C	S150	n/a	n/a	n/a	n/a
102	7511735	1322983H1	SNP00066658	167	472	C	C	A	P150	n/a	n/a	n/a	n/a
102	7511735	1323537H1	SNP00036076	155	254	A	G	A	K77	n/a	n/a	n/a	n/a
102	7511735	1325515H1	SNP00036077	83	288	T	T	C	S89	n/a	n/a	n/a	n/a
103	7511729	1447607R1	SNP00033403	215	589	C	C	A	noncoding	n/a	n/a	n/a	n/a
103	7511729	1447607R1	SNP00033404	39	765	A	A	G	noncoding	n/a	n/a	n/a	n/a
104	7511255	008439H1	SNP00029274	131	1619	C	C	T	noncoding	n/a	n/a	n/a	n/a
104	7511255	1352765H1	SNP00100340	87	2199	T	T	C	noncoding	n/d	n/a	n/a	n/a
104	7511255	1562960H1	SNP00006117	121	1515	C	C	T	noncoding	n/d	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
104	7511255	2188195H1	SNP00100339	205	264	C	C	A	N70	n/a	n/a	n/a	n/a
104	7511255	2188195H1	SNP00143193	237	296	G	G	A	G81	n/a	n/a	n/a	n/a
104	7511255	2531187T6	SNP00100340	128	2260	T	T	C	noncoding	n/d	n/a	n/a	n/a
104	7511255	3111711H1	SNP00006116	52	108	G	G	A	Q18	0.12	n/a	n/a	n/a
104	7511255	5044090H1	SNP00128778	30	87	G	G	A	E11	n/a	n/a	n/a	n/a
104	7511255	6445976H1	SNP00049116	403	390	C	C	T	G112	0.49	n/a	n/a	n/a
104	7511255	6477034H1	SNP00125867	109	177	G	G	A	V41	n/a	n/a	n/a	n/a
105	7511628	1237029F1	SNP00127441	551	1170	G	G	A	noncoding	n/a	n/a	n/a	n/a
105	7511628	123794F1	SNP00004149	169	1408	G	G	C	noncoding	n/a	n/a	n/a	n/a
105	7511628	1273556F1	SNP00099716	309	601	A	A	G	noncoding	0.99	n/a	n/a	n/a
105	7511628	1273556T6	SNP00004149	107	1417	C	G	C	noncoding	n/a	n/a	n/a	n/a
105	7511628	1273556T6	SNP00127441	345	1179	G	G	A	noncoding	n/a	n/a	n/a	n/a
105	7511628	1746976F6	SNP00004149	131	1444	C	G	C	noncoding	n/a	n/a	n/a	n/a
105	7511628	1746976T6	SNP00004149	20	1492	C	G	C	noncoding	n/a	n/a	n/a	n/a
105	7511628	1746976T6	SNP00127441	258	1254	G	G	A	noncoding	n/a	n/a	n/a	n/a
105	7511628	1969606T6	SNP00004149	87	1425	C	G	C	noncoding	n/a	n/a	n/a	n/a
105	7511628	1969606T6	SNP00127441	325	1187	G	G	A	noncoding	n/a	n/a	n/a	n/a
105	7511628	370713T6	SNP00004149	103	1409	G	G	C	noncoding	n/a	n/a	n/a	n/a
105	7511628	370713T6	SNP00127441	341	1171	G	G	A	noncoding	n/a	n/a	n/a	n/a
105	7511628	7709637H1	SNP00004149	51	1410	G	G	C	noncoding	n/a	n/a	n/a	n/a
105	7511628	7709637H1	SNP00127441	289	1172	G	G	A	noncoding	n/a	n/a	n/a	n/a
107	7512343	2593286H1	SNP00001723	127	882	A	A	G	P267	0.34	0.24	0.36	0.65
107	7512343	2697792H1	SNP00111346	74	1010	A	G	A	Q310	n/a	n/a	n/a	n/a
107	7512343	583216R6	SNP00001723	88	883	G	A	G	A268	0.34	0.24	0.36	0.65
108	7512357	1455518H1	SNP00116719	78	160	C	C	T	A46	n/a	n/a	n/a	n/a
108	7512357	1615809F6	SNP00004650	210	894	A	A	C	noncoding	0.56	n/a	n/a	n/a
108	7512357	1615809F6	SNP00025496	31	715	G	G	T	noncoding	n/a	n/a	n/a	n/a
108	7512357	1671471H1	SNP00004651	72	1312	T	C	T	noncoding	0.39	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
108	7512357	5394690H1	SNP00025495	36	52	C	C	T	I10	n/a	n/a	n/a	n/a
108	7512357	568451H1	SNP00025494	80	19	G	G	A	noncoding	n/a	n/a	n/a	n/a
108	7512357	7326673H2	SNP00004650	201	901	A	A	C	noncoding	0.56	n/a	n/a	n/a
108	7512357	7326673H2	SNP00025496	380	722	G	G	T	noncoding	n/a	n/a	n/a	n/a
108	7512357	7372448H2	SNP00004650	193	902	C	A	C	noncoding	0.56	n/a	n/a	n/a
108	7512357	7372448H2	SNP00025496	372	723	G	G	T	noncoding	n/a	n/a	n/a	n/a
108	7512357	7610829J1	SNP00004651	289	1316	C	C	T	noncoding	0.39	n/a	n/a	n/a
108	7512357	7622066J1	SNP00004651	408	1291	T	C	T	noncoding	0.39	n/a	n/a	n/a
108	7512357	7656314J1	SNP00004651	166	1313	T	C	T	noncoding	0.39	n/a	n/a	n/a
109	7511046	082268H1	SNP00019438	222	598	A	A	G	noncoding	n/a	n/a	n/a	n/a
109	7511046	1004538H1	SNP00128595	177	680	A	A	G	noncoding	n/a	n/a	n/a	n/a
109	7511046	1004538T1	SNP00019439	147	633	C	C	T	noncoding	0.99	0.98	0.98	0.99
109	7511046	1270038H1	SNP00001568	42	770	T	C	T	noncoding	n/a	n/a	n/a	n/a
109	7511046	1319295H1	SNP00017409	30	787	T	C	T	noncoding	n/d	n/a	n/a	n/a
109	7511046	1782842H1	SNP00145969	78	493	C	C	A	noncoding	n/a	n/a	n/a	n/a
109	7511046	1795540T6	SNP00128595	93	661	A	A	G	noncoding	n/a	n/a	n/a	n/a
109	7511046	2070794H1	SNP00149262	120	117	G	C	G	W35	n/a	n/a	n/a	n/a
109	7511046	5500374F6	SNP00001568	55	767	C	C	T	noncoding	n/a	n/a	n/a	n/a
109	7511046	7449224T2	SNP00019438	140	599	A	A	G	noncoding	n/a	n/a	n/a	n/a
109	7511046	7449224T2	SNP00128595	58	681	A	A	G	noncoding	n/a	n/a	n/a	n/a
111	7511219	288236R6	SNP00043932	183	249	A	A	G	A67	0.83	0.16	0.55	0.73
112	7510933	008439H1	SNP00029274	131	1985	C	C	T	noncoding	n/a	n/a	n/a	n/a
112	7510933	1352765H1	SNP00100340	87	2565	T	T	C	noncoding	n/d	n/a	n/a	n/a
112	7510933	1562960H1	SNP00006117	121	1881	C	C	T	noncoding	n/d	n/a	n/a	n/a
112	7510933	2188195H1	SNP00100339	205	264	C	C	A	N70	n/a	n/a	n/a	n/a
112	7510933	2188195H1	SNP00143193	237	296	G	G	A	G81	n/a	n/a	n/a	n/a
112	7510933	2531187T6	SNP00100340	128	2626	T	T	C	noncoding	n/d	n/a	n/a	n/a
112	7510933	3111711H1	SNP00006116	52	108	G	G	A	Q18	0.12	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
112	7510933	4574220F6	SNP00049116	80	657	T	C	T	noncoding	0.49	n/a	n/a	n/a
112	7510933	5044090H1	SNP00128778	30	87	G	G	A	E11	n/a	n/a	n/a	n/a
112	7510933	6477034H1	SNP00125867	109	177	G	G	A	V41	n/a	n/a	n/a	n/a
114	7511808	2407209H1	SNP00041726	3	2278	T	T	C	noncoding	n/a	n/a	n/a	n/a
114	7511808	3141911H1	SNP00011024	101	112	A	A	G	noncoding	0.95	n/a	n/a	n/a
114	7511808	6872235H1	SNP00053660	67	2558	C	A	C	noncoding	n/a	n/a	n/a	n/a
114	7511808	7713805H1	SNP00041725	289	2219	C	C	T	noncoding	n/a	n/a	n/a	n/a
114	7511808	882012T6	SNP00053660	2	2590	A	A	C	noncoding	n/a	n/a	n/a	n/a
115	7511817	112389F1	SNP00131869	419	1538	T	T	C	noncoding	n/a	n/a	n/a	n/a
115	7511817	1283673H1	SNP00002979	261	1063	A	A	G	noncoding	0.86	n/d	n/d	0.96
115	7511817	2524056H1	SNP00002978	80	213	A	C	A	noncoding	0.86	0.99	0.63	0.84
115	7511817	2753472H1	SNP00131869	32	1522	T	T	C	noncoding	n/a	n/a	n/a	n/a
115	7511817	2813264T6	SNP00131869	345	1555	T	T	C	noncoding	n/a	n/a	n/a	n/a
116	7511832	1560741T6	SNP00019996	309	360	C	C	G	noncoding	n/a	n/a	n/a	n/a
116	7511832	2557489T6	SNP00019996	308	364	C	C	G	noncoding	n/a	n/a	n/a	n/a
116	7511832	2599523T6	SNP00019996	349	336	C	C	G	noncoding	n/a	n/a	n/a	n/a
116	7511832	311922T6	SNP00019996	330	341	C	C	G	noncoding	n/a	n/a	n/a	n/a
119	7512371	1308518R1	SNP00142971	130	809	C	C	T	noncoding	n/a	n/a	n/a	n/a
119	7512371	1568549T6	SNP00142971	41	840	C	C	T	noncoding	n/a	n/a	n/a	n/a
119	7512371	2207655T6	SNP00005913	337	527	A	A	G	N93	n/a	n/a	n/a	n/a
119	7512371	2323448H1	SNP00005913	170	509	A	A	G	R87	n/a	n/a	n/a	n/a
119	7512371	2412857H1	SNP00039645	219	509	G	G	A	G87	0.68	0.93	0.72	0.76
119	7512371	2812452H1	SNP00028899	156	591	C	T	C	noncoding	n/a	n/a	n/a	n/a
119	7512371	2939152F6	SNP00142971	112	829	C	C	T	noncoding	n/a	n/a	n/a	n/a
119	7512371	3235646T6	SNP00098130	93	768	C	C	T	noncoding	n/d	n/d	n/d	n/d
119	7512371	4534222T1	SNP00142971	68	811	C	C	T	noncoding	n/a	n/a	n/a	n/a
120	7512442	417185H1	SNP00031146	115	766	T	T	C	I229	n/a	n/a	n/a	n/a
121	7512311	1367139R1	SNP00069608	297	1621	G	G	A	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
121	7512311	1393988H1	SNP00065558	76	77	T	T	C	F13	n/a	n/a	n/a	n/a
121	7512311	1797962F6	SNP00105519	239	530	C	C	T	L164	n/a	n/a	n/a	n/a
121	7512311	2304036T6	SNP00069608	211	1623	G	G	A	noncoding	n/a	n/a	n/a	n/a
121	7512311	7315070H1	SNP00069608	327	1605	G	G	A	noncoding	n/a	n/a	n/a	n/a
121	7512311	7977530H2	SNP00148799	220	163	T	C	T	A41	n/a	n/a	n/a	n/a
122	7512474	062073H1	SNP00044544	149	173	C	C	G	L18	n/a	n/a	n/a	n/a
122	7512474	062163H1	SNP00073023	52	461	T	C	T	W114	n/a	n/a	n/a	n/a
122	7512474	062163H1	SNP00105757	72	481	C	C	T	Y120	n/d	1	n/d	n/d
122	7512474	063378H1	SNP00035989	174	189	C	C	T	S23	n/a	n/a	n/a	n/a
122	7512474	4628258T6	SNP00031471	15	792	G	G	A	noncoding	n/a	n/a	n/a	n/a
122	7512474	4651674H1	SNP00031470	131	944	G	G	A	noncoding	0.94	0.93	0.85	0.92
122	7512474	4651674H1	SNP00131613	180	993	T	T	C	noncoding	n/a	n/a	n/a	n/a
122	7512474	862676J1	SNP00044544	150	199	C	C	G	N26	n/a	n/a	n/a	n/a
122	7512474	862676J1	SNP00073023	438	487	C	C	T	F122	n/a	n/a	n/a	n/a
122	7512474	862676J1	SNP00105757	458	507	C	C	T	A129	n/d	1	n/d	n/d